EVOLUTIONARY GENETICS OF THE INNATE IMMUNE SYSTEM IN DROSOPHILA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Erin M. Hill-Burns

August 2009

EVOLUTIONARY GENETICS OF THE INNATE IMMUNE SYSTEM IN DROSOPHILA

Erin M. Hill-Burns, Ph. D.

Cornell University 2009

The innate immune system has drawn interest with its deep evolutionary conservation and importance in response to infection both in invertebrates, providing the only reaction to invading microbes, and as a generalized first reaction in vertebrates, activating the adaptive response. Drosophila has become a useful model for innate immunity, due to the powerful genetic, genomic and comparative resources available. Drosophila-based research has uncovered many genes involved in these pathways, and studies of the population genetics, interspecific divergence, and gene duplication patterns have revealed evidence for selection acting on immune genes, with distinct evolutionary pressures inferred to act on different functional groups.

To investigate variation in immune response within a natural *D. melanogaster* population, we assayed polymorphism in X-linked immune genes and tested correlations between these variants and immune phenotypes. This revealed associations in numerous immune genes with differences in immunocompetence, and strikingly, many of these genotypes appear to associate with sex differences in immune phenotype. Beyond this segregating variation, we also examined divergence between species. In order to quantify regulatory divergence, we have assayed expression before and after infection in *D. melanogaster*, *D. simulans*, and interspecific hybrids. If immune regulation has diverged between species, we expect hybrids to display a disrupted immune response, and we do, in fact, observe immune

dysregulation in the hybrids. This was most notable in the downstream pathway components, indicative of propagation of dysregulation throughout the response.

To further dissect the dysregulation evident in hybrids, we have quantified immune phenotypes in hybrids bearing mutant *melanogaster* alleles for genes throughout the innate immune pathways. Resulting animals are forced to employ only the *simulans* allele at the point of the *melanogaster* mutation. We find that a mutation in one gene in particular, *Dredd*, has a detrimental effect in a hybrid background, indicating that Dredd is likely involved in interactions that have evolved to be specific, such that combinations of *melanogaster* + *simulans* alleles in Dredd interactions lead to a breakdown in immune activation. Overall, we find strong evidence for immune response divergence between species, but the degree of functionally relevant divergence appears to be greatest among signaling molecules.

BIOGRAPHICAL SKETCH

Born and raised in Delaware, Erin developed an interest in Biology, and more specifically, Genetics, virtually as soon as she was introduced to the topics in science classes. Eager to pursue an education toward performing genetics research, she entered into science-related summer internships at the Delaware Public Health Laboratory after her Junior and Senior years of high school. While much of her time during these summers consisted of data entry, she was exposed to laboratory techniques and equipment, getting a helpful introduction into the laboratory environment.

Erin attended the University of Delaware from 1999 to 2003, where she received a Bachelors of Science in Biology, with a concentration in Cellular and Molecular Biology and Genetics, and a minor in Chemistry. Spending some of her time as a teaching assistant in an Introductory Biology lab, she also became involved in an undergraduate research project in the lab of Dr. David Usher, under the guidance of Dr. Robert Hodson. This allowed her to further develop her interest in genetics research by sequencing the Apolipoprotein A-II gene in the red-eared slider turtle, for comparative sequence studies across species. During a winter session break, Erin changed gears and took a month-long research internship in the lab of Dr. Leemor Joshua-Tor at Cold Spring Harbor Laboratory, studying the crystallization of a DNA repair protein, MutY. While providing additional interesting lab experience, this confirmed that she was not especially interested in the further pursuit of focused biochemical research.

While her undergraduate career did not provide an in-depth background in the fields of population genetics and evolution, she had gained enough exposure to be intrigued by the field, and she decided to focus her graduate research in this area. Upon graduation from college in 2003, Erin arrived at Cornell University to pursue a Ph.D. in Genetics. Receiving a Cornell Presidential Genomics Fellowship, Erin spent

her first year doing rotation projects in the labs of Dr. Andrew Clark, Dr. Charles Aquadro, and Dr. William Provine. While each of these allowed her to explore various areas of evolutionary genetics and genomics (and even the history of genetics), she ultimately joined the lab of Dr. Andrew Clark, to pursue research into the genetics of the innate immune system of Drosophila and its evolution between species.

ACKNOWLEDGMENTS

While the research described here includes little officially collaborative work, it was certainly not done in isolation – the help and support of many others have made this process a more productive one, made the products stronger, made the research go more smoothly, and made the whole journey more pleasurable. Firstly, I have to thank my advisor, Dr. Andrew Clark. Andy's guidance through my entire graduate experience has strengthened my research practices and output, challenging me along the way. His drive is infectious, and his consistently positive outlook has made for an enjoyable work environment; it has been extremely rewarding to be a part of his lab group.

I also have to thank the members of my thesis committee, Dr. Charles Aquadro, Dr. Richard Harrison, and Dr. Brian Lazzaro. They have always been ready with helpful input and advice, eager to be available for whatever assistance I might have needed.

Other colleagues in the lab and outside have helped in numerous ways. Tim Sackton and Kurt McKean provided input on the Drosophila immune response and more, and they frequently assisted with experiments. Xiaoyun Wang provided technical assistance for general lab needs, as well as sequencing for some preliminary analysis for my work. Many others pitched in for bits of time throughout my larger experiments; some of these investigations would not have been possible without their help. I have to particularly acknowledge the assistance of undergraduate lab assistants Anna Beavis, Ryan Cummings, and Tracy Mak, who spent a large portion of one summer helping me infect flies and assay bacterial load phenotypes. I also want to thank Amanda Larracuente, for her help with a million little things around lab – from tips on writing computer scripts to ordering lab supplies – and for being a great friend along the way.

Outside of lab, friends in my department and outside of it have made the past six years a lot of fun. Through gatherings, camping trips, happy hours and more, I truly appreciate all of the friends who have been there for me in any number of ways. My family has also been incredibly supportive throughout my entire educational journey. Even when not entirely understanding why I might want to spend my days and nights studying fruit flies, my parents and brother, grandparents, aunts, uncles, cousins, and more have always been there for me, eager to hear how things have been going.

Finally, I am incredibly grateful to my wonderful husband, Andrew. He has been amazingly supportive over the years, encouraging me along the way. I particularly appreciate all of his help and support in the recent months, during my push to finish this thesis. From doing laundry and cooking to reading drafts of papers and discussing research problems and troubleshooting, he has been there for me in every way possible.

TABLE OF CONTENTS

Biographical Sketch..... iii
Acknowledgements..... v
Table of Contents..... vii
List of Figures..... ix
List of Tables..... x
List of Abbreviations..... xi
Chapter 1. Introduction..... 1
Introduction..... 1
References..... 6
Chapter 2. X-linked Variation in Immune Response in *Drosophila melanogaster*..... 9
Abstract..... 9
Introduction..... 9
Materials and Methods..... 13
Construction of Lines..... 13
Genotyping of SNPs Across Lines..... 14
Bacterial Cultures and Infections..... 16
Bacterial Load Quantification..... 16
TaqMan RT-PCR..... 17
Statistics and Association Testing..... 18
Results..... 24
Variation Observed in X-Linked Immune Genes..... 24
Genetic Variation in Bacterial Load..... 25
Genetic Variation in Immune Gene Induction..... 37
Correlations Between Induction and Load Phenotypes..... 38
Discussion..... 41
References..... 46
Chapter 3. Functional Regulatory Divergence of the Innate Immune System in Interspecific Drosophila Hybrids..... 50
Abstract..... 50
Introduction..... 50
Materials and Methods..... 53
Fly Lines and Crosses..... 53
Bacterial Cultures and Infections..... 54
Transcript Quantification Using BeadChip Arrays..... 54
Bacterial Load Quantification..... 55
Statistical Analysis..... 55
Results..... 58
Expression Changes After Infection..... 58
Patterns of Non-Additivity in F1 Hybrid Expression..... 61
Expression Levels Among Different Functional Classes of Immune Genes in F1 Hybrids..... 63
Coordinated Regulatory Divergence in Hybrid Expression Change after Infection 65

Bacterial Clearing Ability in Hybrids and Parental Species. 65
Discussion..... 68
References..... 74
Chapter 4. Genome-wide Regulatory Divergence After Bacterial Infection in Interspecific Drosophila Hybrids..... 78
Abstract..... 78
Introduction..... 78
Materials and Methods..... 81
Fly Lines and Crosses..... 81
Bacterial Cultures and Infections..... 81
Transcript Quantification via Short-Read Sequencing..... 81
Alignment of Reads to Reference Genomes..... 82
Quantification and Normalization of Read Depth..... 82
Allele-specific Read Density in Hybrids..... 83
Statistical Analysis..... 84
Results..... 84
Transcripts Identified by Short-Read Sequencing..... 84
Transcript Abundance in Parental and Hybrid Flies..... 85
Functional Differences in Hybrid Dysregulation..... 89
Allele-specific Difference in Transcript Abundance in Hybrids..... 91
Discussion..... 97
References..... 103
Chapter 5. Disruption of the Innate Immune Response in Mutation-Perturbed Interspecific Drosophila Hybrids..... 107
Abstract..... 107
Introduction..... 107
Materials and Methods..... 110
Fly Lines and Crosses..... 110
Bacterial Cultures and Infections..... 110
Bacterial Load Quantification..... 113
Quantification of Transcript Abundance using Pyrosequencing..... 113
Results..... 114
Effect of Genotype on Bacterial Clearing Ability..... 114
Effect of Genotype on Transcription Levels..... 116
Dredd Genotype Has No Effect on Relative Transcript of Parental Alleles..... 119
Discussion..... 119
References..... 123
Appendix A. Supplementary Information for Chapter 2..... 126
Appendix B. Supplementary Information for Chapter 3..... 165
Appendix C. Supplementary Information for Chapter 4..... 171
Appendix D. Supplementary Information for Chapter 5..... 172

LIST OF FIGURES

2.1 Line Means of Bacterial Load After Infection 26
2.2 Sex Difference in Mean Bacterial Load 28
2.3 Genes in Drosophila Immune-Related Pathways 30
2.4 Example Effects of SNP on Bacterial Load 32
2.5 Epistatic Interactions Associating with Bacterial Load After Infection 36
2.6 Line Means for Bacterial Load vs. Tsf1 Induction Levels 40
3.1 Distribution of P-values for Tests of Expression Differences 60
3.2 Proportion of Immune Genes Differently Expressed 62
3.3 F1 vs. Parental Mean Expression and Induction/Repression Levels 64
3.4 Genes Throughout Toll and imd, Clustered Based on Changes in Expression 66
3.5 Genes Uniquely Induced in F1 Hybrids 67
3.6 Bacterial Load after Infection in Parental and Hybrid Flies 69
4.1 Counts of Genes Transcribed 86
4.2 Genome-wide Transcript Levels in Hybrids and Parents 87
4.3 Changes in Transcript Level After Infection in Hybrids and Parental Species 88
4.4 Expression Levels Before and After Bacterial Infection 90
4.5 Average Proportions of *D. melanogaster* Allele in Expressed Genes 95
4.6 Differences in Proportions of the Two Parental Alleles in Hybrids 99
5.1 Cross Scheme Used for Mutant Hybrids and Control Crosses 112
5.2 Differences in Bacterial Load after Infection 115
5.3 Bacterial Load After Infection in Flies Bearing Wild Type or Mutant Alleles 117
5.4 Transcript Levels of *Dpt* 118
5.5 Proportion of *Dpt* Transcript from the *D. melanogaster* Parental Allele 120
A.1 SNPlex Oligos 133

LIST OF TABLES

2.1 Genes Selected for Genotyping 15

2.2 SNPs Associating with Load Phenotypes 31

2.3 Multiple SNP Clusters Associating with Load Phenotypes 35

2.4 SNPs Associating with Immune Gene Induction Phenotypes 39

4.1 Pathways Showing Significant Effects of Infection on Expression 92

4.2 Pathways Showing Significant Effects of Infection on Allelic Proportions 96

5.1 Genes Disrupted in Exelixis P-element Insertion Lines 111

A.1 Autosomal Genotyping Results 126

A.2 Resequencing Primers 129

A.3 Genotypes at Each SNP 144

A.4 TaqMan Primer and Probe Sequences 164

B.1 Genes Included on BeadChips 165

C.1 Transcribed Genes Classified among Functional Groups 171

D.1 Mutant D. melanogaster Stocks Used in Crosses 172

D.2 Primer and Probe Sequences for *Dpt* Pyrosequencing 172

LIST OF ABBREVIATIONS

Aminosugars = Aminosugars metabolism pathway

AMP = antimicrobial peptide

ANOVA = analysis of variance

bp = base pair

cDNA = complementary DNA

cfu = colony-forming units

CT = critical threshold

D = Tajima's D statistic

d.f. = degrees of freedom

D. melanogaster = *Drosophila melanogaster*

D. simulans = *Drosophila simulans*

DNA = deoxyribonucleic acid

E. faecalis = *Enterococcus faecalis*

F = F-statistic

FDR = false discovery rate

Glycan deg. = Glycan structures – degradation pathway

Glycophos. = Glycophospholipid metabolism pathway

K_A/K_S = ratio of non-synonymous to synonymous substitutions

kb = kilobase

KEGG = Kyoto Encyclopedia of Genes and Genomes

LB = Luria-Bertani

LD = linkage disequilibrium

mRNA = messenger RNA

n = number of lines/samples

NF-κB = nuclear factor κB

OD = optical density

oligo = oligonucleotide

oligo-dT = oligonucleotide comprised of a string of T bases

π = measure of sequence variation

P. rettgeri = *Providencia rettgeri*

P-value = probability value

PCR = polymerase chain reaction

PGRP = peptidoglycan recognition protein

Purine = Purine metabolism pathway

q-value = false discover rate for probability value

r^2 = coefficient of determination

rDNA = ribosomal DNA

RNA = ribonucleic acid

RNA-seq = short-read sequencing of transcript

RT-PCR = reverse transcriptase polymerase chain reaction

S = number of segregating sites

S.E. = standard error of the mean

S. marcescens = *Serratia marcescens*

SNARE = SNARE interactions in vesicular transport pathway

SNP = single nucleotide polymorphism

SR = scavenger receptor

Std. Err. = standard error of the mean

τ = Spearman's τ, measure of correlation

t = t-test statistic for differences in means

θ_W = Watterson's theta

χ^2 = Chi-squared test

CHAPTER 1

INTRODUCTION

The innate immune pathways have been found to be highly conserved among taxa throughout the animal kingdom, comprising the only response to infection in invertebrates and providing a critical activation step of the adaptive immune pathways in vertebrates (MEDZHITOV and JANEWAY 1997). Due to the evolutionary importance of this system, it has been investigated in a variety of organisms. The Drosophila immune response has become a model of particular interest, with extensive resources available for genetic and comparative analyses. In recent years, many genetic studies in Drosophila have clarified the genetic factors and their interactions in the innate immune response (BRENNAN and ANDERSON 2004; FERRANDON *et al.* 2007; IRVING *et al.* 2004).

The most studied portion of Drosophila immunity, the humoral response, includes the Toll and imd pathways, which are homologous to those in the mammalian innate immune response (KIMBRELL and BEUTLER 2001). These pathways involve the recognition of invading microbes, signal transduction after activation, and transcription of antimicrobial peptides (AMPs) and other effector molecules as attack mechanisms. The Drosophila immune response also includes cellular mechanisms, including melanization and encapsulation or phagocytosis of invading cells, as well as less well-defined interactions with other pathways involved in stress response.

Despite the constancy of the genes that comprise the main functions of these pathways in Drosophila and other species, they have been shown to evolve under the influence of positive selection at higher levels than those seen for non-immune genes (SCHLENKE and BEGUN 2003). Furthermore, genes involved in different functions within the immune response appear to show different levels of polymorphism and divergence within and between species. In some studies examining levels of variation

in subsets of immune genes, antimicrobial peptides have been found to have low levels of polymorphism and little evidence of positive selection (CLARK and WANG 1997; DATE *et al.* 1998; LAZZARO and CLARK 2003; RAMOS-ONSINS and AGUADE 1998), whereas some recognition proteins, including scavenger receptors, have been found to evolve under positive selection (LAZZARO 2005). In a recent study of sequence divergence between multiple Drosophila species in genes throughout the immune system, more evidence was found for distinct patterns of selection acting on genes belonging to different functional groups. Here, (SACKTON *et al.* 2007) found greater signs of positive selection acting on genes encoding recognition proteins than with genes involved in either signaling or effector functions.

Beyond examinations of sequence-level differences, other investigations have quantified variation in immune-related genes in natural populations as well as the functional consequences of such variation (LAZZARO *et al.* 2006; LAZZARO *et al.* 2004). These experiments have surveyed flies in natural populations to determine which autosomal genes harbor variation that can impact phenotype and therefore be acted on by selection. These studies uncovered polymorphisms in multiple genes throughout the immune pathways that associate with immunocompetence phenotypes, indicating that a variety of potential targets for selection may exist in natural populations.

In order to more fully explore the effects of naturally occurring polymorphisms on immune response phenotypes, we have assayed variation in X-linked immune genes and quantified associations between these genotypes and response to infection, described in Chapter 2. Here, we report that a variety of X-linked genes contain naturally segregating polymorphisms that correlate with bacterial load phenotypes after infection, consistent with previous studies, but we have also found that a substantial number of these variants associate with sex differences in the immune

response. This presents evidence for sexual dimorphism existing among X-linked genetic variants, potentially allowing for these to be maintained as segregating alleles within a population.

Beyond investigations of population level variation in the immune response, studies of divergence in immune phenotypes between species can provide an additional perspective on the evolutionary pressures acting upon genes in these pathways. To this end, we have examined the divergence in regulation of the innate immune response between Drosophila species by quantifying levels of transcription in immune-related genes before and after infection in *D. melanogaster*, *D. simulans*, and their F1 hybrids.

Since interspecific F1 hybrids contain genomes of two diverged species in a single individual, they have the potential to display disrupted phenotypes as a consequence of incompatibilities between parental alleles (LANDRY *et al.* 2007; TRUE and HAAG 2001). Previous investigations have revealed hybrid dysregulation of developmental and enzyme expression phenotypes (DICKINSON *et al.* 1984; PARKER *et al.* 1985; WHITT *et al.* 1977), and genome-wide studies have found widespread patterns of non-additive expression levels in interspecific hybrids (HEGARTY *et al.* 2005; MICHALAK and NOOR 2003; MOEHRING *et al.* 2007; RANZ *et al.* 2004; RENAULT *et al.* 2009).

If regulation of the immune response has diverged between *D. melanogaster* and *D. simulans*, we expect that the F1 hybrids of these species should display dysregulatory immune phenotypes, reflective of the divergence between the parental species. To investigate this, we quantified expression levels of immune genes in *D. melanogaster*, *D. simulans*, and their F1 hybrids before and after infection using custom Illumina BeadChips, described in Chapter 3. With this, we do find evidence for dysregulatory patterns of expression in the F1 hybrids in response to infection.

Despite overall trends toward additivity in expression levels after infection in the hybrids, numerous genes appear as outliers from these trends. Furthermore, genes belonging to recognition, signaling, and effector functional groups show distinct patterns of dysregulation in the hybrids, providing further evidence for dissimilar regulatory divergence among these groups of genes.

To examine broader influences of dysregulation in hybrids in response to bacterial infection, we also quantified transcript abundance in *D. melanogaster*, *D. simulans*, and their F1 hybrids before and after infection using Illumina short-read sequencing of cDNA sequences (RNA-seq), explained in Chapter 4. Through this, we were able to quantify genome-wide expression patterns in these flies, revealing dysregulation in non-immune pathways, as well as in the immune response genes in response to infection. Furthermore, we were able to quantify allele-specific transcript levels within the hybrid flies, revealing skewed patterns of expression between the two parental alleles before and after infection. There was some tendency for genes downstream of a gene with skewed allelic expression to be similarly skewed, but this effect was weak.

While regulatory differences in F1 hybrids can reveal divergent portions of the immune response in the parental species, it does have the limitation of detecting only interactions with effects dominant to the regulatory phenotypes of the conspecific alleles, since genes in the F1 hybrids may still be able to maintain normal interactions with alleles from both parents present for each gene in a complex. In order to further perturb the immune system, we have examined hybrids bearing mutant *D. melanogaster* alleles for genes throughout the innate immune pathways, described in Chapter 5. With a dysfunctional allele from one parent, genes involved in critical interactions in the immune response may be less able to interact if the genes involved have diverged between the parental species such that interspecific interactions are less

functional. This investigation pinpointed the protein Dredd, critically important for imd pathway function, as one in which interactions have evolved between species, yielding dysfunctional effects of interspecific complexes.

Overall, these studies have clarified patterns of variation in innate immune response within and between species of Drosophila. Despite largely conserved structure and function of the immune response, there is increasing evidence for divergence in gene sequences and regulation in this system. While variation segregating in natural populations that associates with immune response phenotypes and sex differences may exist within immune genes, selection to fix diversity between species appears to act distinctly on different functional groups of immune genes. Even in the face of dysregulatory patterns of expression, though, the hybrids show a generally functional immune response, indicating a robustness of the immune response to some perturbations.

REFERENCES

BRENNAN, C. A., and K. V. ANDERSON, 2004 Drosophila: the genetics of innate immune recognition and response. Annu Rev Immunol **22:** 457-483.

CLARK, A. G., and L. WANG, 1997 Molecular population genetics of Drosophila immune system genes. Genetics **147:** 713-724.

DATE, A., Y. SATTA, N. TAKAHATA and S. I. CHIGUSA, 1998 Evolutionary history and mechanism of the Drosophila cecropin gene family. Immunogenetics **47:** 417-429.

DICKINSON, W. J., R. G. ROWAN and M. D. BRENNAN, 1984 Regulatory Gene Evolution: Adaptive Differences of Alcohol Dehydrogenase in *Drosophila melanogaster* and *Drosophila simulans*. J. Hered. **52:** 215-225.

FERRANDON, D., J. L. IMLER, C. HETRU and J. A. HOFFMANN, 2007 The Drosophila systemic immune response: sensing and signalling during bacterial and fungal infections. Nat Rev Immunol **7:** 862-874.

HEGARTY, M. J., J. M. JONES, I. D. WILSON, G. L. BARKER, J. A. COGHILL *et al.*, 2005 Development of anonymous cDNA microarrays to study changes to the Senecio floral transcriptome during hybrid speciation. Mol Ecol **14:** 2493-2510.

IRVING, P., L. TROXLER and C. HETRU, 2004 Is innate enough? The innate immune response in Drosophila. C R Biol **327:** 557-570.

KIMBRELL, D. A., and B. BEUTLER, 2001 The evolution and genetics of innate immunity. Nat Rev Genet **2:** 256-267.

LANDRY, C. R., D. L. HARTL and J. M. RANZ, 2007 Genome clashes in hybrids: insights from gene expression. Heredity **99:** 483-493.

LAZZARO, B. P., 2005 Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans*. Genetics **169:** 2023-2034.

LAZZARO, B. P., and A. G. CLARK, 2003 Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. Mol Biol Evol **20:** 914-923.

LAZZARO, B. P., T. B. SACKTON and A. G. CLARK, 2006 Genetic variation in *Drosophila melanogaster* resistance to infection: a comparison across bacteria. Genetics **174:** 1539-1554.

LAZZARO, B. P., B. K. SCEURMAN and A. G. CLARK, 2004 Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. Science **303:** 1873-1876.

MEDZHITOV, R., and C. A. JANEWAY, JR., 1997 Innate immunity: impact on the adaptive immune response. Curr Opin Immunol **9:** 4-9.

MICHALAK, P., and M. A. NOOR, 2003 Genome-wide patterns of expression in Drosophila pure species and hybrid males. Mol Biol Evol **20:** 1070-1076.

MOEHRING, A. J., K. C. TEETER and M. A. NOOR, 2007 Genome-wide patterns of expression in Drosophila pure species and hybrid males. II. Examination of multiple-species hybridizations, platforms, and life cycle stages. Mol Biol Evol **24:** 137-145.

PARKER, H. R., D. P. PHILIPP and G. S. WHITT, 1985 Gene regulatory divergence among species estimated by altered developmental patterns in interspecific hybrids. Mol Biol Evol **2:** 217-250.

RAMOS-ONSINS, S., and M. AGUADE, 1998 Molecular evolution of the *Cecropin* multigene family in Drosophila. functional genes vs. pseudogenes. Genetics **150:** 157-171.

RANZ, J. M., K. NAMGYAL, G. GIBSON and D. L. HARTL, 2004 Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. Genome Res **14:** 373-379.

RENAULT, S., A. W. NOLTE and L. BERNATCHEZ, 2009 Gene Expression Divergence and Hybrid Misexpression between Lake Whitefish Species Pairs (*Coregonus* spp. Salmonidae). Mol Biol Evol **26:** 925-936.

SACKTON, T. B., B. P. LAZZARO, T. A. SCHLENKE, J. D. EVANS, D. HULTMARK *et al.*, 2007 Dynamic evolution of the innate immune system in Drosophila. Nat Genet **39:** 1461-1468.

SCHLENKE, T. A., and D. J. BEGUN, 2003 Natural selection drives Drosophila immune system evolution. Genetics **164:** 1471-1480.

TRUE, J. R., and E. S. HAAG, 2001 Developmental system drift and flexibility in evolutionary trajectories. Evol Dev **3:** 109-119.

WHITT, G. S., D. P. PHILIPP and W. F. CHILDERS, 1977 Aberrant Gene Expression during the Development of Hybrid Sunfishes (Perciformes, Teleostei). Differentiation **9:** 97-109.

CHAPTER 2

X-LINKED VARIATION IN IMMUNE RESPONSE IN *DROSOPHILA MELANOGASTER*

Abstract

A fundamental challenge of immune systems is to defend against unpredictably diverse pathogens. Understanding an immune system's evolutionary robustness must begin with analysis of genetic variation in its components, and here we examine effects of X-linked genetic variation on immunocompetence of Drosophila. We generated 168 lines of *D. melanogaster* with X chromosomes extracted from a natural population into a co-isogenic autosomal background and genotyped the lines at 88 SNPs in 20 X-linked immune genes. We tested the genotypic variation among lines for association with bacterial load phenotypes after infection with Gram-positive *Enterococcus faecalis* and Gram-negative *Serratia marcescens*. We also quantified immune gene expression in a subset of lines to evaluate associations with induction phenotypes. We find that polymorphisms in many genes correlate with load and/or expression phenotypes. These results are comparable to those from experiments testing effects of naturally occurring autosomal polymorphisms on immune variation. Many of the associations found here, however, are sex-specific or sexually antagonistic. This supports the theory that sexually antagonistic variation may be maintained on the X chromosome even in the presence of selection that typically purges variation in these genes facing regular hemizygosity.

Introduction

The deep evolutionary conservation of many specific genes in innate immunity underscores the potent forces of natural selection maintaining this vital function.

While it is widely accepted as the ancestral form of immune response, its role in the activation of adaptive immune response further motivates investigation into variation in its function (MEDZHITOV and JANEWAY 1997). Drosophila has been used as a valuable model organism to identify and characterize functions of the components of innate immune pathways as well as the evolutionary patterns present among the genes of these pathways (reviewed in BRENNAN and ANDERSON 2004; FERRANDON *et al.* 2007; IRVING *et al.* 2004). The humoral response, resulting in the production of antimicrobial peptides in response to bacterial or fungal infection, relies mainly on Toll and imd signal transduction pathways, both of which are homologous to pathways in mammalian immunity (reviewed in KIMBRELL and BEUTLER 2001). The cellular component, on the other hand, incorporates both phagocytic engulfment as well as melanization and encapsulation of infecting particles. While less well defined in the Drosophila model, portions of other systems also appear to impact the effectiveness of immune response, including JAK/STAT and JNK signaling pathways, hematopoesis and iron metabolism.

Population genetic analysis can be used to determine whether sequence polymorphism and divergence patterns among Drosophila genes in innate immune pathways are consistent with signatures of selection acting within and between species of flies. If, for example, the innate immune pathways are involved in an evolutionary "arms race" with pathogenic organisms, genes in these pathways would be expected to show signs of positive selection driven by evolutionary pressure to counter virulence mechanisms of invading microbes. When signs of selection (as inferred by sequence comparisons within *Drosophila simulans* populations and between *D. simulans* and *D. melanogaster*) in immune genes and non-immune genes were evaluated, immune genes as a group were found to have higher K_A/K_S ratios than non-immune genes, providing evidence for elevated adaptive evolution (SCHLENKE and BEGUN 2003).

Since receptor, effector and signaling proteins function in different portions of the immune response pathways, these may be exposed to differing levels of contact with invading microbes and may display non-uniform levels of functional redundancy or pleiotropy. Thus, genes from different functional groups may be exposed to distinct selective pressures. Antimicrobial peptides, which might be expected to encounter unique selective pressures due to their direct interactions with invading microbes, have shown little sign of positive selection, bearing low levels of amino acid divergence (CLARK and WANG 1997; DATE *et al.* 1998; LAZZARO and CLARK 2003; RAMOS-ONSINS and AGUADE 1998). Furthermore, sequence analyses of immune-related receptors have shown evidence for purifying selection in peptidoglycan recognition proteins (PGRPs), while others, including some scavenger receptors (SRs), appear to be rapidly evolving under pressures consistent with positive selection (JIGGINS and HURST 2003; LAZZARO 2005). On a deeper evolutionary timescale, sequence comparisons between immune genes in multiple Drosophila species (based on full-genome sequence data) have shown striking differences among functional groups of immune genes, with recognition molecules showing much more positive selection than either signaling or effector genes (SACKTON *et al.* 2007).

Beyond using sequence data and the analysis of polymorphism and divergence to infer levels and modes of selection that have previously acted on immune genes (both individually or in groups), other studies have investigated correlations between autosomal variation in genotype and immune response phenotype in naturally occurring populations of Drosophila (LAZZARO *et al.* 2006; LAZZARO *et al.* 2004; SACKTON *et al*., *submitted*). These experiments tested associations between naturally occurring genetic variation in immune-related genes and post-infection bacterial load. In these studies, genetic variation in many of the immune genes was found to associate significantly with one or more of the bacterial load phenotypes. Specifically, many

polymorphisms in autosomal genes encoding recognition and signaling proteins (but not antimicrobial peptides) associate consistently with bacterial load phenotypes, suggesting that not all functional classes of immune-related genes harbor equally influential genetic variation.

The focus of this study is X-linked immune genes, which may be under unique regulatory and selective pressures simply because they are hemizygous in males, are dosage compensated, and face elevated influence of random genetic drift due to their smaller effective population size. As a consequence, the X chromosome should favor the more rapid fixation of beneficial recessive alleles and more rapid loss of harmful recessive alleles compared to autosomes (CHARLESWORTH *et al.* 1987; SINGH *et al.* 2008). Thus, with different selective pressures compared to autosomal genes, X-linked immunity genes are expected to maintain different standing levels of variation, and segregating polymorphisms in these genes may have different impacts on phenotype.

Different exposures of X-linked genes to selection in males and females can also contribute to sexual dimorphism. (RICE 1984) suggested that X-linked sexually antagonistic alleles may more freely impact sexually dimorphic traits than can those on autosomes. In fact, the X chromosome appears to favor the maintenance of sexually antagonistic variation (GIBSON *et al.* 2002). The presence of sexually antagonistic selection among X-linked genes may be able to preserve polymorphism; if a given allele is slightly deleterious in one sex, it may be maintained in the population by being beneficial to the other sex. Immune-related genes may be particularly prone to displaying sexual dimorphism in Drosophila, since males and females have been shown to have different evolutionary optima for energetic expenditure on immune response, and thus their respective immune responses may differ based on other conditions such as food or reproductive resource availability (MCKEAN and NUNNEY 2001; MCKEAN and NUNNEY 2005). If sexually antagonistic traits are responsible for

some of the observed sexual dimorphism, variation in X-linked genes could contribute to phenotypic differences, and so X-linked variation in immune genes could face distinct selective pressures.

In this report we investigate the standing levels of variation in X-linked immune genes in natural populations of *D. melanogaster* and quantify the impacts of that variation on immune response phenotypes. We genotyped 168 lines at single nucleotide polymorphisms (SNPs) across 20 X-linked immunity loci and quantified post-infection bacterial load and immune gene expression phenotypes. We found significant variation across the lines for bacterial load after infection, and we were able to identify polymorphisms in immune-related genes that associate with immune response phenotypes individually and in interacting pairs of SNPs. Additionally, some of the genetic variation was found to associate with a sex difference in immune competence, with alleles acting either in a sex-specific or sexually antagonistic manner. This provides evidence for X-linked genetic variation in immune-related loci associating with both phenotypic variation among lines and sex differences in these phenotypes.

Materials and Methods

Construction of Lines: *D. melanogaster* females were collected from apple orchards near Ithaca, NY by Todd Schlenke and Brian Lazzaro in 2004. Isofemale lines were established and kept under laboratory conditions for fewer than five generations prior to isogenization. X chromosomes were isogenized in these lines, by individually mating males from each line to females of the highly inbred balancer stock *FM7a*, B^{1} sc^{8} v^{Of} w^{a} y^{31d}. From each of these crosses, three female offspring were individually mated to *FM7a* males. Since the balancer chromosome bears the co-dominant marker *Bar*, heterozygous female offspring could be selected for the crosses

each generation. The crossing scheme was repeated for each line in triplicate for a total of seven generations to replace the background autosomes from the natural population. This resulted in 168 lines, each homozygous (or hemizygous) for a unique X chromosome from nature and all co-isogenic for the replaced autosomes. The degree of background replacement was quantified by subsequent SNP genotyping, finding concordance between the marker background and the isogenized lines in 99.6% of all assays (1191 tests out of 1196 examined; see Table A.1 for full autosomal genotyping results).

Genotyping of SNPs Across Lines: Candidate immune-related genes were selected for genotyping based on their previously indicated connections to immune responses in genetic studies and/or large-scale expression assays (Table 2.1). These genes include well-characterized members of the Toll and imd pathways, as well as genes involved in other aspects of the response to infection, including hematopoesis and iron metabolism. There is a significant over-representation of the genes in the JAK/STAT pathway among the X-linked immunity genes (χ^2, $P = 7.7 \times 10^{-5}$), and several of these genes were included in our study. Notably, none of the 20 genes encoding antimicrobial peptides genome-wide exist on the X chromosome in *D. melanogaster*, so our investigation lacks any genotyped members of this class of immune genes.

To identify SNPs for genotyping, the entire gene region for each gene, including roughly 1 kb upstream and downstream, was resequenced in eight of the X-extraction lines (see Table A.2 for list of primers used). Table 2.1 reports summary statistics for these sequence alignments, calculated using DnaSP (ROZAS and ROZAS 1995), except for Tajima's D, which was calculated with VariScan so as not to exclude all sites with missing data (VILELLA *et al*. 2005). Once polymorphism data were collected for all genes, SNPs were chosen for genotyping from among those

Table 2.1 Genes Selected for Genotyping

Functional Group	Gene Name	Cytological Position	Sequence Length	n	S	π	θ_W	D	SNPs
Recognition	*PGRP-LE*	13F1	1027	8	4	0.0014	0.0015	-0.2218	1
	PGRP-SA	10C6	1414	8	0	0	0	NC	0
Signal	*domeless*	18D13-E1	1484	8	7	0.0028	0.0025	0.3364	6
Transduction	*Dredd*	1B12-13	2452	8	13	0.0018	0.0021	-1.3748	4
	hemipterous	11D10	5444	8	65	0.0067	0.0073	-1.1266	14
	hopscotch	10B5-6	5388	8	41	0.0040	0.0040	-0.5257	7
	pole hole	3A1	3834	8	47	0.0059	0.0058	-0.8160	9
	Tak1	19D2	6318	8	115	0.0086	0.0083	-0.0373	5
	Traf2	7D16	3327	8	70	0.0092	0.0091	0.3354	3
	Traf3	14C4	2704	8	31	0.0058	0.0063	-1.3593	2
Other	*Dsor1*	8D2-3	1171	6	6	0.0043	0.0043	NC	2
	lozenge	8D5-6	3452	8	44	0.0066	0.0069	0.9657	4
	multi sex combs	8D2	2738	8	21	0.0043	0.0040	-1.3101	2
	Ntf2	19E7	2969	8	52	0.0078	0.0082	-0.7046	2
	outstretched	17A5	2953	8	13	0.0026	0.0028	0.2036	1
	Pvf1	17E1-6	5372	8	110	0.0118	0.0117	NC	10
	Rps6	7C2	2278	8	54	0.0103	0.0106	0.3699	3
	Ser7	9A2	2339	8	15	0.0023	0.0027	-0.0835	3
	Transferrin 1	17A9	3135	8	28	0.0034	0.0037	-1.4206	6
	unpaired 2	17A3	2524	8	21	0.0043	0.0039	-0.4479	3
	unpaired 3	17A4	5261	8	190	0.0164	0.0186	-0.8004	4

n = number of lines sequenced, S = segregating sites, D = Tajima's D (NC = not calculated), SNPs = number genotyped

present at relatively intermediate frequencies in the samples and spaced approximately 500-1,000 bp apart within the genes. SNPs in high linkage disequilibrium (LD) with one another were generally avoided. Wherever possible, nonsynonymous SNPs were included; however, the selection of SNPs genotyped included those from exonic, intronic, 5' and 3' untranslated and intergenic regions. In total, 91 SNPs were chosen from among these 20 genes for genotyping across all 168 lines. *PGRP-SA* was not included in the SNP genotyping due to a complete absence of detectable variation found in the resequenced sample.

To identify the genotype for each line at each selected SNP, the SNPlex system (Applied Biosystems, Foster City, CA) was used. Oligos were designed and synthesized to query the genotype of all 91 SNPs (see Figure A.1 for oligo and SNP information). The associated GeneMapper software was used to make the initial SNPlex allele calls, and these were followed by manual inspection. 88 of the 91 SNP assays in the SNPlex system yielded useful genotypic information across the 168 lines (see Table A.3 for genotype calls at each site for all lines).

Bacterial Cultures and Infections: Bacterial stocks were chosen based on previous use for immune challenges in *D. melanogaster*. The strain of Gram-positive bacterium *Enterococcus faecalis* was derived from that used by LAZZARO *et al*. (2006) (identified via 16S rDNA sequence and results of API 20Strep substrate utilization testing). We also selected Gram-negative *Serratia marcescens,* derived from ATCC strain 13880, which also had been used in previous studies (LAZZARO *et al.* 2006; LAZZARO *et al.* 2004). Bacterial cultures for infections were grown from freezer stocks, and cultures were grown overnight at 37° to a final concentration of $OD_{600} \approx 1.0$ for each day of infections.

Bacterial Load Quantification: Bacterial clearing ability of the lines was measured through quantification of bacterial load after infection with bacteria,

following LAZZARO *et al.* (2004). *D. melanogaster* were individually infected by pricking their thoraces with 0.1-mm tungsten needles (Fine Science Tools, Foster City, CA) dipped in bacterial culture. For each bacterium, a block design of infections was used: each round of infections was repeated three times over six days in a two-week span, with half the lines infected on a given day. For each round of infections, 12 males and 12 females from each line, aged approximately 3-10 days, were infected (for technical feasibility, several people served as infectors on each day, but lines were randomized among infectors from day to day). Approximately 26-30 hours after infection, three groups of three flies per line were homogenized in 500 μl of LB broth, and were then plated onto LB agar plates using a spiral plater (Spiral Biotech, Bethesda, MD). Homogenates with *E. faecalis* bacteria were diluted 1:1000 pre-plating to achieve a countable level of colonies. Plates were kept at either room temperature or at 37° to allow bacterial colonies to grow until they could be counted by a colony counter. These counts allowed inference of the concentration of bacteria in each homogenate sample. Plates were visually inspected to confirm that colonies counted were consistent with size and morphology expected. Thus, for each line, both sexes were infected with each of two bacteria, and each round of infections included three replicates for each sex and bacterial infection of every line, over three rounds of infections. This yielded nine independent biological replicates of each infection with a total of over 21,000 flies infected and 5,046 plates counted.

TaqMan RT-PCR: In addition to bacterial load, expression phenotypes were measured after infection for a subset of 16 lines, chosen from the phenotypic tails of sex difference in load after infection with *E. faecalis*. For each of these lines, 30 males and 30 females were infected with *E. faecalis* injections. Eight hours after infection, three replicates of each line and sex, most with 8-10 flies each, were snap-frozen in liquid nitrogen, along with three replicates of uninfected flies. RNA was extracted

using a Trizol:chloroform protocol. cDNA was then synthesized from the isolated nucleic acid and diluted to fulfill TaqMan protocol requirements (Applied Biosystems, Foster City, CA). Transcripts were quantified using TaqMan RT-PCR, including antimicrobial peptide genes (*DiptericinA*, *Defensin*, and *Metchnikowin*), along with X-linked immune-related genes (*Peptidoglycan Recognition Protein-SA* and *Transferrin1*) and ribosomal protein *RpL32* as an autosomal reference gene (see Table A.4 for probe and primer sequence information). We measured the CT value for each sample (number of PCR cycles at which the level of fluorescence for the sample crosses a constant critical threshold value) and used the reciprocal, 1/CT, as a proxy for expression for further calculations.

Statistics and Association Testing: Bacterial load was determined for each sample in terms of colony-forming units per fly (cfu/fly). Estimates of bacterial density from Drosophila homogenates (pools of three flies each) range from 1.0×10^0 to 4.0×10^6 cfu (corresponding to 0.3×10^0 to 1.3×10^6 cfu/fly). All empty plates were recorded as true zero counts (on a log scale) rather than as missing data; all plates with density calculated above 3.0×10^6 were too dense to be accurately counted, so these were all assigned to have densities of 4.0×10^6, which probably underestimates the number of colonies in most cases. Residuals from the analysis of variance on the raw cfu counts were distributed non-normally, and log transformation yielded an adequate fit of residuals to the normal distribution. Statistical analyses were carried out using the R software (R Development Core Team, 2007) and SAS/STAT software with the SAS system (SAS Institute, Inc., Cary, NC). To test for significant effect of Line, as well as a Line × Sex interaction on variation in bacterial load (for each bacterial infection), we used the mixed models:

$$y_{ijklmn} = \mu + \text{Line}_i + \text{Sex}_j + (\text{Line} \times \text{Sex})_k + \text{Day}_l + \text{Infector}_m + \text{Plater}_n + \varepsilon_{ijklmno} \qquad (1a)$$

$$y_{ijklm} = \mu + Line_i + Sex_j + Day_l + Infector_k + Plater_m + \varepsilon_{ijklmn} \quad (1b)$$

$$y_{ijkl} = \mu + Sex_i + Day_j + Infector_k + Plater_l + \varepsilon_{ijklm} \quad (1c)$$

where y is equal to ln(cfu/fly) (bacterial count), Line (i = 1...168), Sex (j = 1,2), and the Line × Sex interaction are fixed effects, and Day (k = 1...6), Infector (l = 1...6) and Plater (m = 1...2) are treated as random effects using the R package lme4. ε is the error term. The full model (1a) was compared to the partially reduced model (1b) using ANOVAs to test for the effect of a Line × Sex interaction term. Similarly, the partially reduced model (1b) was compared to the reduced model (1c) to determine the effect of Line differences. To test the significance of each effect, load phenotypes were permuted 1,000 times in R (for each bacterium), while keeping line, sex, and random effects constant. The coefficients of the model tests from these permutated data provided a null distribution as a basis of comparison for the actual Line and Line × Sex effects estimated from the data, and *P*-values were calculated for each. The proportions of variance explained by models incorporating just Line effects or Line × Sex interaction effects (r^2) were also calculated for each bacterial infection using R.

Mixed models were also employed to test for associations between line genotypes and the above phenotypes. Here, differences in each load phenotype (*e.g.* total *E. faecalis* load) between two alleles for each SNP were tested for significance using mixed models:

$$y_{ijklmno} = \mu + Allele_i + Sex_j + (Allele \times Sex)_k + Line_l + Day_m + Infector_n + Plater_o + \varepsilon_{ijklmnop} \quad (2)$$

where y represents the phenotype of interest, Allele (i = 1,2) corresponds to the genotype at the SNP in question, as a fixed effect, and Sex (j = 1,2) is also included as a fixed effect. The Allele × Sex interaction term was included as a fixed effect to quantify the effects of sex on SNP associations with bacterial load. Line (l = 1...168), Day (m = 1...6), Infector (n = 1...6), and Plater (o = 1,2) are all included as random effects. Nearly identical models were used to test allelic effects on bacterial load in either males or females individually, with the exception that these did not include Sex as a fixed effect:

$$y_{ijklm} = \mu + Allele_i + Line_j + Day_k + Infector_l + Plater_m + \varepsilon_{ijklmn} \quad (3)$$

Because of the potential for linkage disequilibrium among SNPs, tests of association were not all independent, so significance was assessed using permutation tests. Each SNP was tested individually, and genotypes were permuted 1,000 times in R, relative to load phenotypes and the line, sex, day, infector, and plater values. The resulting coefficients for Allele or Allele × Sex effects provided a null distribution against which to compare the coefficients from tests with actual values, providing *P*-values for each. False discovery rate (FDR) was estimated by calculating q-values for each test using the qvalue R package (STOREY 2002). To determine the proportion of variance explained by each SNP, r^2 values for models including each SNP alone as a fixed effect were calculated using R.

Associations between SNP genotypes and expression phenotypes were also examined using the mixed model:

$$y_{ijklmnopq} = \mu + Allele_i + Sex_j + Treatment_k + (Allele \times Treatment)_l + (Allele \times Treatment \times Sex)_m + RpL32_n + Line_o + Plate_p + Replicate_q + \varepsilon_{ijklmnopqr} \quad (4)$$

where y represents the expression level of the gene of interest (1/CT). Here, the fixed effects include Allele (i = 1...168), corresponding to the genotype at a given SNP, Treatment (j = 1,2), representing "infected" or "uninfected" state of the flies, Allele × Treatment, the interaction of SNP genotype and infection state to test for the effect of change in expression level after infection, Allele × Treatment × Sex, the influence of sex on this induction effect, and RpL32, the expression of *RpL32,* as a covariate to normalize the expression phenotype measured. Line (o = 1...16), Plate (p = 1,2), and Replicate (q = 1...3) were all included in the model as random effects. For each phenotype, every SNP was tested individually. As above, phenotype-genotype combinations were permuted 1,000 times in R, and the coefficients for SNP × Treatment and SNP × Treatment × Sex effects provided null distributions against which to compare the actual coefficients and assign *P*-values. FDR values were again estimated using qvalue. r^2 values for models including each SNP alone as a fixed effect were calculated using R.

Haplotypes of SNPs were assessed for the presence of blocks of high LD across the X chromosome using the program Haploview (BARRETT *et al.* 2005). Since these lines are homozygous for any variation, these comparisons essentially involve counts of gametes. Missing SNP data (Table A.3) were imputed using the program fastPHASE version 1.1 (SCHEET and STEPHENS 2006). The 10 haplotype blocks (sets of 2-3 SNPs within 9 different genes) indicated to have significant levels of LD by the Haploview program were tested for associations with both load and expression

phenotypes. These association tests were performed in the same manner as listed above for single SNPs, using the mixed models:

$$y_{ijklmno} = \mu + \text{Haplotype}_i + \text{Sex}_j + (\text{Haplotype} \times \text{Sex})_k + \text{Line}_l + \text{Day}_m + \text{Infector}_n + \text{Plater}_o + \varepsilon_{ijklmnop} \quad (5a)$$

$$y_{ijklm} = \mu + \text{Haplotype}_i + \text{Line}_j + \text{Day}_k + \text{Infector}_l + \text{Plater}_m + \varepsilon_{ijklmn} \quad (5b)$$

where Equation 5a tests for genotypic effect of Haplotype (i = 1...10) on bacterial load and Haplotype × Sex effects on bacterial load in all flies; Equation 5b was used to test for effects of Haplotype on bacterial load in males or females individually. As above, *P*-values were assigned based on null distributions of coefficients of haplotype effects from permuted data sets.

In addition to associations between single SNP genotypes and phenotypes, effects of epistatic interactions were also examined. Here, the effects of interactions between every possible combination of SNP pairs (both within and between genes) were tested. Rigorous inference of pairwise epistasis normally requires consideration of all nine two-locus genotypes, and the usual caveats of fitting linear models with sparse marginal counts apply (COCKERHAM 1954). Here we have homozygous lines, so there are only four genotypes to contrast, and only one degree of freedom for tests of the single epistatic component and fixed marginal frequencies, so the model is closer to that of (CHEVERUD and ROUTMAN 1995). With the low-frequency alleles of some SNPs, not every SNP pair allowed for valid tests of associations with all four genotype combinations, so these pairs were not included. For each valid test, two-way ANOVAs were performed to test associations with each phenotype using models both with and without SNP interaction terms; a significant difference between the fit of the

two models to the data indicated an effect of the SNP interaction. The full and reduced models compared here are as follows:

$$\text{full:} \quad y_{ijk} = \mu + SNP1_i + SNP2_j + (SNP1 \times SNP2)_k + \varepsilon_{ijkl} \tag{6a}$$

$$\text{reduced:} \quad y_{ij} = \mu + SNP1_i + SNP2_j + \varepsilon_{ijk} \tag{6b}$$

where y is the load or expression phenotype, SNP1 (i = 1...88) and SNP2 (j = 1...88) are the two SNPs of interest, and SNP1 × SNP2 is the interaction term of the allelic effects of these two SNPs. Due to the computational time needed to test all SNP combinations, these simpler linear models were applied, using estimated line means for the load and expression phenotypes. To accommodate the same random effects as above, the phenotypic values used were the least squares means for each line obtained using mixed models in SAS, based on Equation 1b for load phenotypes and Equation 4 for expression phenotypes. These SNP interaction effects were tested for associations with load in males, females, and both sexes combined, along with the sex difference in load (female load - male load) for each bacterium. In addition, associations were tested with induction of expression (infected - uninfected expression levels) in males, females and both combined. As above, with these ANOVA tests, we calculated *P*-values by permuting the genotype-phenotype combinations 1,000 times and comparing actual F statistics to the null distributions of F statistics from tests with the permuted data. Again, r^2 values were calculated to quantify the proportion of variance explained by the interaction term; this was determined from the difference in r^2 values of the full and reduced models.

Beyond tests of association between the genotypes and phenotypes of these lines, we also tested the ability of expression phenotypes to predict load after

infection. More specifically, we tested the effects of uninfected expression levels and induction of expression on *E. faecalis* levels after infection. These tests used the following models:

$$y_{ijklmn} = \mu + Exp_i + Sex_j + (Exp \times Sex)_k + RpL32_l + Plate_m + Rep_n + \varepsilon_{ijklmno} \quad (7a)$$

$$y_{ijklmnop} = \mu + Exp_i + Infection_j + (Exp \times Infection)_k + Sex_l + (Exp \times Infection \times Sex)_m + RpL32_n + Plate_o + Rep_p + \varepsilon_{ijklmnopq} \quad (7b)$$

Here, y is the load phenotype, ln(cfu per fly), and the model includes Exp (the expression level of the gene assayed, 1/CT) as a fixed effect, along with RpL32 (expression level of *RpL32*) as a covariate to normalize the expression level of the gene of interest. Sex (j = 1,2), along with sex interaction terms are also included as fixed effects in both (except when each sex is considered individually). Plate (m = 1,2) and Rep (n = 1...3, replicate) were also included as random effects in these models. In Formula 7a, Infection status is not included; here only uninfected or infected samples are considered at one time. In Formula 7b, however, the Expression × Infection term accounts for induction effects (if uninfected and infected flies have significantly different expression levels). Again, significance values for these tests were calculated based on null distributions of coefficients from tests using data permuted 1,000 times.

Results

Variation Observed in X-Linked Immune Genes: To quantify effects of naturally occurring X-linked variation in immune genes on immune phenotypes, X chromosomes from a natural population of *D. melanogaster* were extracted into co-isogenic autosomal backgrounds. To find polymorphic sites in the immune genes in

these lines, 21 candidate genes were resequenced in eight sample lines. In the approximately 67.5 kb of sequence obtained in these lines (including intronic, exonic, 3' and 5' untranslated and intergenic regions), 947 SNPs were uncovered, one SNP about every 71 bases on average. Out of the SNPs found in this sample, 172 are in coding regions, and 23 of these (13%) are nonsynonymous. An analysis of the sequence polymorphisms seen here shows non-skewed values of Tajima's D, but somewhat lower levels of variation (Table 2.1) than have been seen in other population genetic analyses of Drosophila immune genes. Compared to studies including autosomal immune and non-immune genes from non-African populations of *D. melanogaster* (ANDOLFATTO 2001; RAMOS-ONSINS and AGUADE 1998), these X-linked immune genes have much lower values of θ_W than the autosomal immune genes (*t*-test, $P = 0.0002$), while still showing significantly higher levels of variation than X-linked non-immune genes (*t*-test, $P = 0.0127$). Most of the autosomal immune genes assayed for polymorphism, though, have been AMPs, and since none of these exist on the X chromosome for comparison, this disparity in levels of variation between X-linked and autosomal immune genes could be due at least in part to differences among functional groups.

Genetic Variation in Bacterial Load: We calculated bacterial load means for each *D. melanogaster* X-extraction line 26-30 hours after infection with *E. faecalis* and with *S. marcescens*. The line means in load span a range of 9.82 to 16.73 natural log colony-forming units per fly (ln(cfu/fly) for *E. faecalis* and 7.23 to 10.29 ln(cfu/fly) for *S. marcescens*, representing 1,007-fold and 21-fold ranges which span 1.6 to 1.1 average within-line phenotypic standard deviations, respectively (Figure 2.1)). Analyses of variance showed that this variation was significant among lines for both bacteria ($P < 0.001$ for each). Furthermore, the line means of load for the two bacterial species are not correlated (correlation coefficient = 0.035, N.S., Figure 2.1C).

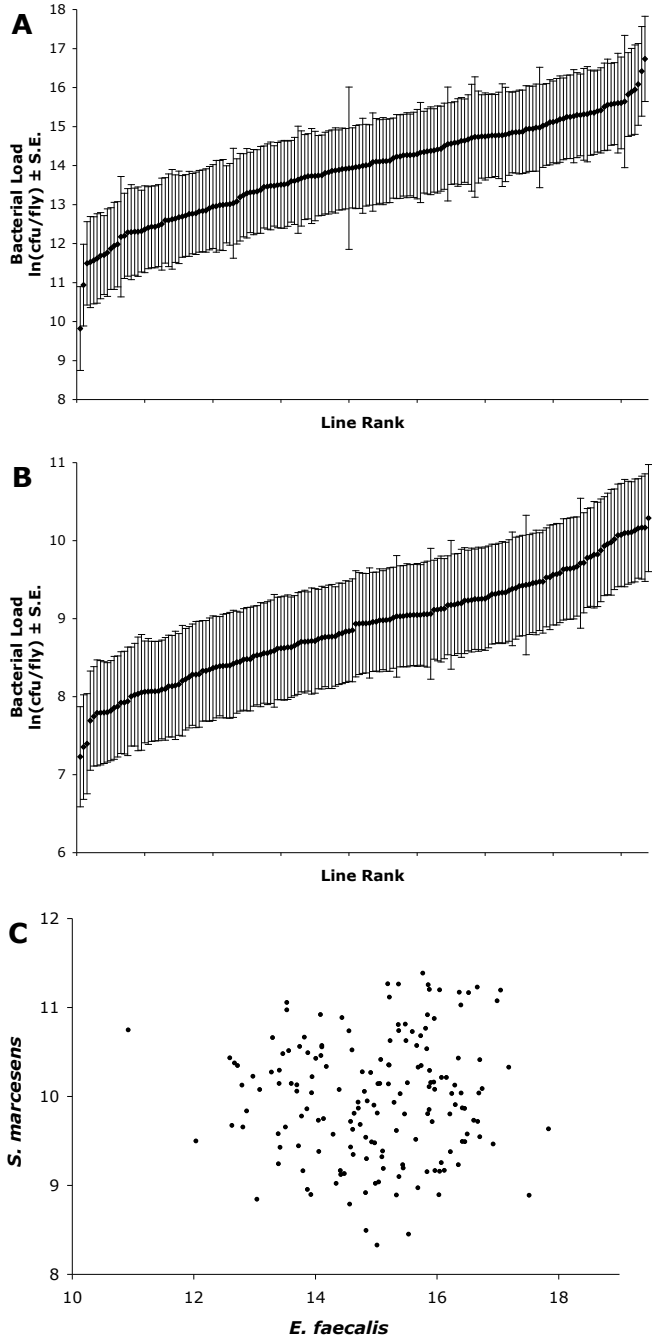


Figure 2.1 Line means of bacterial load after infection with A) *Enterococcus faecalis* or B) *Serratia marcescens* and C) scatterplot of means of load for the two bacteria. Lines are plotted in rank order for each bacterium in A and B. Bacterial load is measured as the natural log of the count of colony-forming units per fly ln(cfu/fly), shown with the standard errors of the mean.

A lack of correlation of load across bacterial types has been noted in earlier studies (LAZZARO *et al.* 2006; LAZZARO *et al.* 2004), and the interpretation has been that bacterial-host interactions are bacterial species-specific, which can lead to different immune response dynamics, depending on the virulence mechanisms employed by the bacteria and the host response to this infection.

In addition to differences among lines in bacterial load after infection, we also find variation among lines in differences in load between males and females. Figure 2.2 shows the sex differences in mean load of both bacteria (in terms of ln(cfu/fly) across all the lines). These differences (female mean cfu/fly – male mean cfu/fly) range from 1.9×10^7 to -3.6×10^6 cfu/fly (from 1080-fold higher in females to 31-fold higher in males) for *E. faecalis*, with a median difference (across the line means) of 6.4×10^6 cfu/fly. No effort was made to control for body size between sexes, but these sex differences on load are much larger than what might be expected from body size differences alone. For *S. marcescens*, the differences range from 9.3×10^4 to -2.8×10^4 cfu/fly (from 11-fold higher in females to 12-fold higher in males), with a median difference of 5.3×10^3 cfu/fly. Significantly more than half the lines bear mean differences greater than zero for both *E. faecalis* (χ^2, d.f.= 1, $P = 2.3 \times 10^{-18}$) and *S. marcescens* loads (χ^2, d.f. = 1, $P = 0.0055$). With most lines here displaying higher bacterial load in females than in males after infection, this could imply that males in these lines have more effective immune responses than females. MCKEAN and NUNNEY (2005) find the opposite effect (higher load in males after infection) with plentiful food and mates, yet this study also highlights the condition-dependent nature of these results. Furthermore, these experiments have included load assays after different types of bacterial infections, which might not be expected to yield the same levels of bacterial load or sex differences in load.

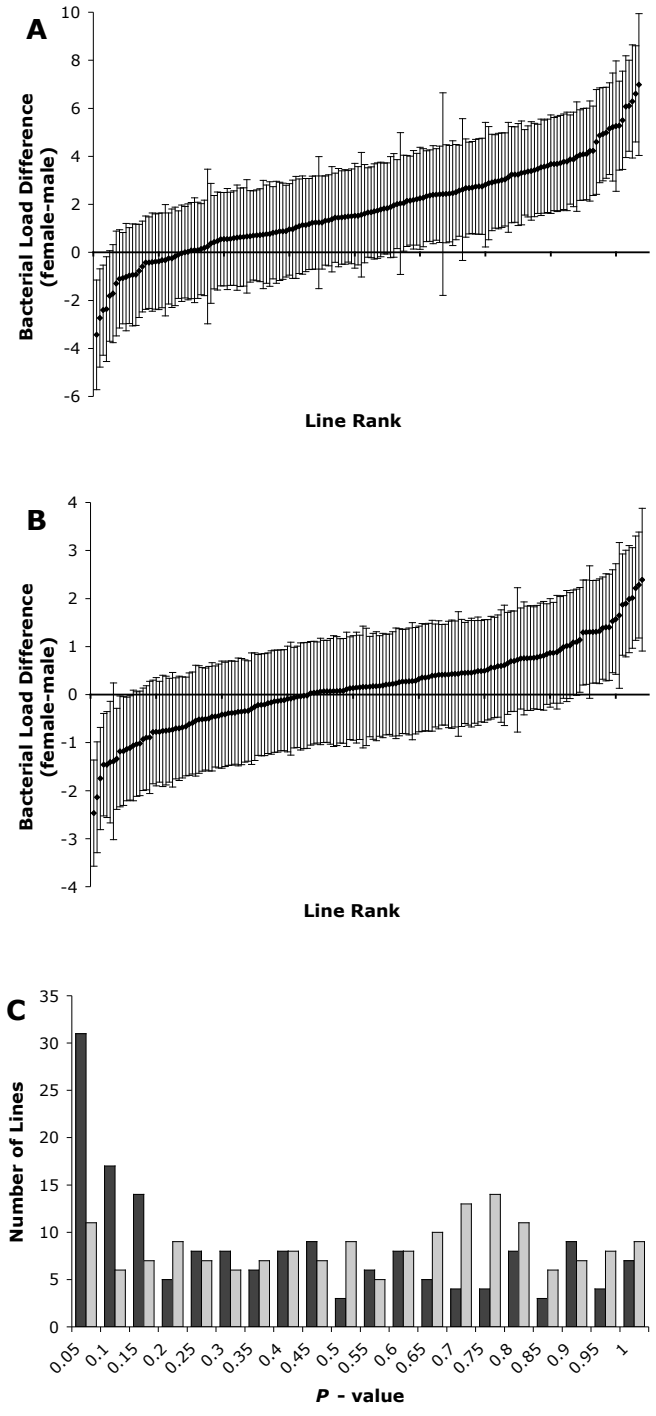


Figure 2.2 Sex differences in mean bacterial load after infection with (A) *E. faecalis* or (B) *S. marcescens*, displayed as (female ln(cfu/fly) – male ln(cfu/fly)) ± (standard error of the difference) and (C) histogram of *P*-values of *t*-tests of sex difference in all lines after *E. faecalis* infection (black bars) or *S. marcescens* infection (gray bars).

Many of these lines show significant differences between male and female load, particularly when infected by *E. faecalis*. When each line is tested for sex effect on load, the distribution of *P*-values is highly skewed from an expectation of equal load in both sexes, with an excess of *t*-tests with $P < 0.05$ (χ^2, d.f. = 1, $P = 8.8 \times 10^{-16}$) in flies infected with *E. faecalis*; however, in those infected with *S. marcescens, P*-values from *t*-tests of sex effects show no significant departure from the expected distribution (χ^2, d.f. = 1, $P = 0.36$) (Figure 2.2C). Even though these lines show a wide range of differences between sexes, bacterial load after infection in males and that in females are significantly correlated for both *E. faecalis* and *S. marcescens* (Spearman's τ, $P = 0.0025$, $P = 3.21 \times 10^{-11}$). Thus, for most of the lines, higher (or lower) bacterial load remains relatively consistent in both sexes. As expected from the greater sex differences in lines infected with *E. faecalis*, though, load values in males and females infected with this bacterium are less strongly correlated than are those in flies infected with *S. marcescens*.

Genotypic variation among the extraction lines was tested for association with variation observed in immune phenotypes. X-linked genes from immune-related pathways (Figure 2.3) were chosen as candidates and genotyped to determine standing levels of variation. 88 SNPs in 20 candidate immune genes (Table 2.1) were individually tested for allelic effects on bacterial load phenotypes after infection with both *E. faecalis* and *S. marcescens*. Table 2.2 lists the q-values, based on *P*-values calculated from permuted null distributions, for those SNPs that showed at least one phenotypic association with FDR < 10% ($q < 0.1$). While these tests reveal possible associations with multiple SNPs within different immune genes, any given SNP typically explains less than 8% of the variance in bacterial load phenotypes (Table 2.2).

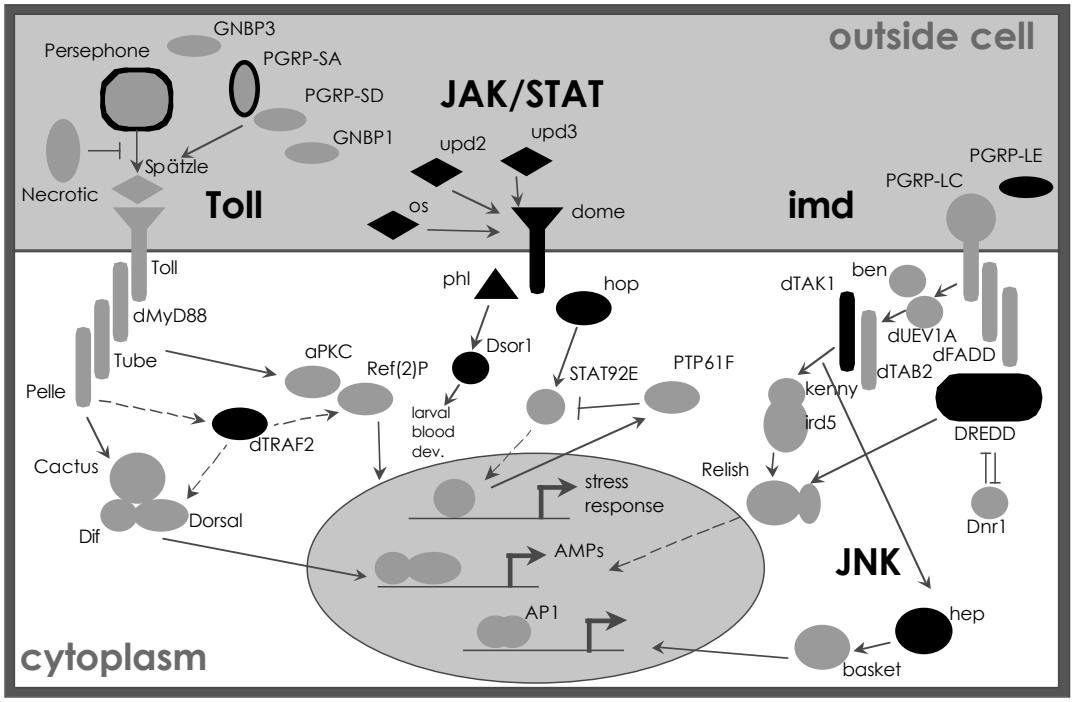


Figure 2.3 Genes in Drosophila immune-related pathways. Those colored black are X-linked genes included in this study, those outlined in black are X-linked, but were not genotyped here. Pathway genes and interactions included based on information in previous studies (ARBOUZOVA and ZEIDLER 2006; FERRANDON *et al.* 2007; FOLEY and O'FARRELL 2004; LECLERC and REICHHART 2004; STRONACH and PERRIMON 2002; WASSARMAN *et al.* 1995).

Table 2.2 SNPs Associating with Load Phenotypes

Functional Class	Gene	Location	Change	Ef female	Ef male	Ef ALL	Sm female	Sm male	Sm ALL	Ef SNP*Sex	Sm SNP*Sex
Signal Transduction	*hopscotch*	exon	**V->L**	0.688	0.739	0.981	<0.001*** 0.7	0.360	0.032* 0.3	0.172	0.946
	hopscotch	exon		<0.001*** 0.4	0.025* 0.5	0.981	<0.001*** 0.8	0.076* 0.2	<0.001*** 0.4	<0.001*** 6.2	0.022* 0.9
	hopscotch	exon		<0.001*** 0.4	0.120	<0.001*** 0.2	<0.001*** 0.1	0.735	0.372	0.274	0.043* 0.7
	hemipterous	exon	**Y->C**	0.476	<0.001*** 1.0	<0.001*** 0.6	0.445	0.801	0.372	0.323	0.498
	hemipterous	exon	**A->S**	0.476	0.215	0.981	0.847	0.520	0.758	0.029* 5.9	0.997
	hemipterous	exon		0.258	0.973	0.981	0.732	0.123	0.168	0.029* 6.2	0.997
	Tak1	intron		0.258	0.973	0.404	0.775	<0.001*** 0.8	<0.001*** 0.5	0.304	0.417
	TRAF3	exon		0.999	<0.001*** 1.0	<0.001*** 0.0	0.283	0.393	0.042* 0.1	0.000*** 6.2	0.997
Secreted	*outstretched*	exon	**A->S**	0.404	0.973	0.521	0.117	<0.001*** 0.8	<0.001*** 0.4	0.850	<0.001*** 1.1
	upd2	exon		0.999	0.356	0.536	0.067* 0.4	0.633	0.140	0.925	0.731
	upd3	intergenic		0.306	<0.001*** 0.4	<0.001*** 0.2	<0.001*** 0.4	0.336	<0.001*** 0.2	0.123	<0.001*** 0.9
Iron Metabolism	*Tsf1*	exon		0.999	<0.001*** 0.9	0.012* 0.4	0.912	0.814	0.906	0.151	0.997
Hematopoesis	*lozenge*	exon		0.306	<0.001*** 1.0	0.688	0.092* 0.6	<0.001*** 1.2	<0.001*** 0.8	<0.001*** 7.7	0.997
	Pvf1	5' UTR		0.999	0.973	0.981	0.851	0.135	0.443	0.987	0.037* 0.6
	Pvf1	exon		0.043* 1.5	0.120	<0.001*** 0.9	0.832	0.633	0.890	0.987	0.731
	Pvf1	exon		0.000*** 0.5	0.221	<0.001*** 0.3	0.065* 0.3	0.829	0.271	0.487	0.034* 0.7
	Pvf1	intron		0.453	0.226	0.134	0.445	0.109	0.028* 0.0	0.808	0.997
	Rps6	exon		0.999	0.951	0.981	0.820	0.000*** 0.6	0.017* 0.2	0.987	0.037* 1.0
Serine Protease	*Ser7*	intergenic		0.999	<0.001*** 0.2	0.022* 0.1	0.117	<0.001*** 0.4	0.833	0.029* 6.2	<0.001*** 0.9

Q-values shown for SNPs that associate with at least one load phenotype (based on having FDR q < 0.1). Percent of total phenotypic variance explained shown for each SNP with q < 0.1 (below q-value). Change = amino acid change associated with SNP, where applicable. 69 of 88 SNPs show no association with load phenotypes with q < 0.1. *q < 0.1, ***q < 0.001.

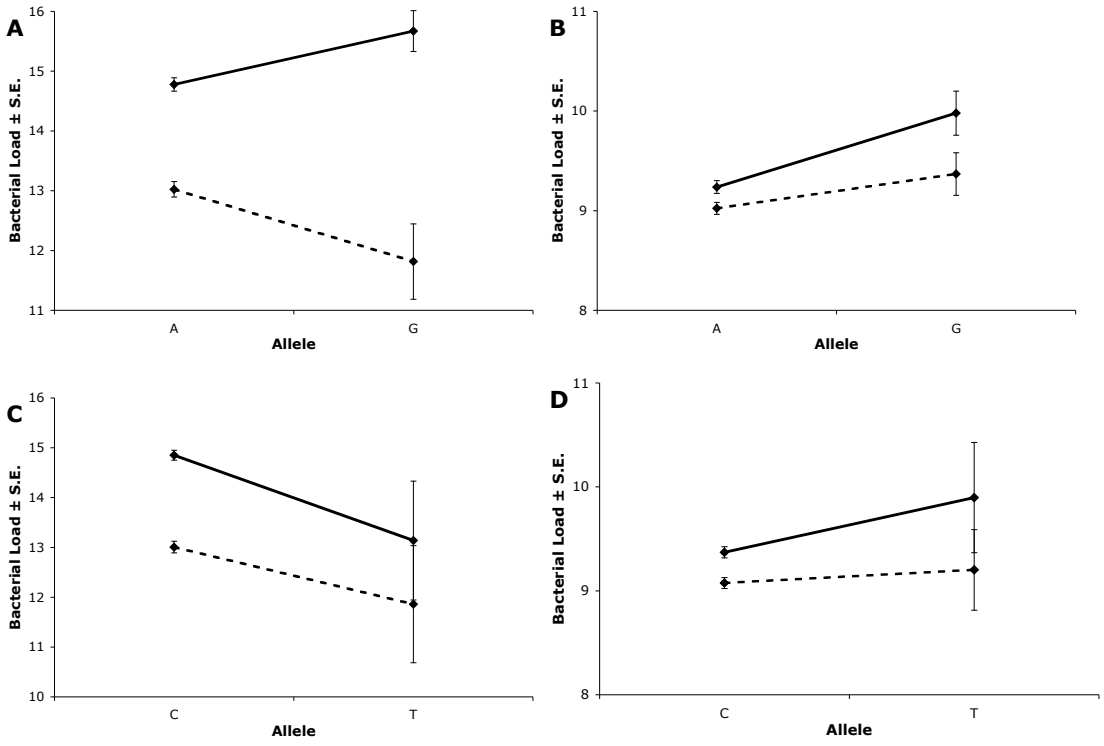


Figure 2.4 Example effects of SNP on both *E. faecalis* (A & C) and *S. marcescens* (B & D) load after infection in females (solid lines) and males (dashed lines). SNPs shown include *hop* exon7-01 (synonymous, residue 870) in A & B and *hop* exon7-02 (synonymous, residue 968) in C & D.

Out of the 19 SNPs in 12 genes that associate with one or more of the load phenotypes at this level, eight associate (at least marginally) with phenotypes for both bacteria; however, five of these show opposite effects across bacteria in one or both sexes. Examples of this include two SNPs in the gene *hopscotch* (*hop*). As depicted in Figure 2.4, one SNP in exon 7 of the gene (panels A & B) has significant allelic effects on load in both males and females, with both bacteria. The effects of the two infections in males, though, appear in opposite directions - a substitution from the "A" allele to the "G" allele of this SNP associates with a lower *E. faecalis* load, yet a higher *S. marcescens* load after infection. Similarly, for the second SNP in *hop* exon 7 (306 bp downstream from the first, panels C & D), allelic effects are once again significantly associated with load in females infected by either bacterium, yet the load variation in females occurs in opposite directions for the two bacteria.

Besides the distinct phenotypes and associations appearing in response to each of the two bacterial infections, some SNPs also associate with load variation in a sex-specific or even sexually antagonistic manner. Out of the 19 SNPs associating with load, 12 show evidence of sex interactions influencing the associations with the load of one or both bacterial infection. More specifically, several of these actually appear to have opposite effects in males and females; most of these SNPs do not show significant associations in both sexes individually, though, lessening our ability to find clear instances of sexually antagonistic associations. Any potentially sexually antagonistic effects, though, only appear with one of the two bacteria in each case. An example of this is seen with the SNP in *hop* Exon 7 in Figure 2.4A.

In addition to single-SNP tests of association, we consider the possibility that multiple SNPs that fall into particular haplotype configurations might correlate with differences in immune function. We identified haplotypes as collections of SNPs in LD and subsequently tested these for associations with immune phenotypes.

Haplotype blocks with significant levels of LD (as defined by the program Haploview, see Materials and Methods) were identified in 10 sets of 2-3 SNPs across nine genes. Most of these blocks involved fairly closely located SNPs. Less than 17% of all SNP pairs within 1 kb of each other were found to be in high LD, consistent with previous findings that LD decays quickly along the Drosophila genome (CARBONE *et al.* 2006; LONG *et al.* 1998). These blocks of high LD were tested for associations with load phenotypes. Variation in six of these haplotype blocks significantly associated ($P < 0.05$) with differences in one or more bacterial load phenotype (Table 2.3). Twelve individual haplotype-phenotype associations appear with $P < 0.05$, 10 of which have a FDR < 10% ($q < 0.1$). Many of these associations only appear with *E. faecalis* load phenotypes; three haplotypes appear to associate with *S. marcescens* load - but only in sex × haplotype interactions. Additionally, for all the haplotype blocks that show significant associations, none of the SNPs included in each haplotype associate individually with the same load phenotype (with FDR < 10%). Furthermore, many of the SNPs included in these haplotypes are noncoding or synonymous; the only nonsynonymous SNPs among these clusters are the two in the *Pvf1* haplotype. These two SNPs, both located in Exon 1, appear to be outside of the identified PDGF domain and the putative signal peptide of the gene, so no obvious disruption of function is inferred from their amino acid changes. Overall, it appears that the SNPs within the significantly associating haplotypes are most likely in linkage disequilibrium with any variation that could directly lead to phenotypic differences in these lines.

For each possible pair of SNPs among the 88, we tested for pairwise epistasis based on the significance of the interaction term in two-way ANOVAs. Figure 2.5 depicts those pairs of genes within the Toll, imd, JAK/STAT and JNK pathways that contain SNPs with interactions associating with load phenotypes with q-values less than 0.1 for tests of models using SNP pair interactions to explain variation in

Table 2.3 Multiple SNP Clusters Associating with Load Phenotypes

Gene(SNPs)	Location	Ef female	Ef male	Ef ALL	Sm female	Sm male	Sm ALL	Ef Sex*Hap	Sm Sex*Hap
Dredd (2,3,4)	Exon2, Intron1, 5'intergenic	0.139	0.226	0.489	0.195	0.931	0.443	0.191	0.021*
phl (6,7)	Exon4	0.187	0.250	0.017*	0.650	0.096	0.473	0.722	0.328
mxc (1,2)	5'intergenic & Exon2	0.013*	0.037*	0.101	0.325	0.150	0.123	0.028*	0.285
hep (9,10,11)	Exon5, Exon4, Exon3	0.043*	0.084	0.121	0.311	0.336	0.209	<0.001**	0.026*
Pvf1 (3,4)	Exon1	<0.001**	0.001**	0.019*	0.095	0.274	0.104	0.348	0.625
Tak1 (1,2)	Intron4 & Exon3	0.956	0.945	0.974	0.575	0.486	0.484	0.241	0.033*

* P < 0.05, ** P < 0.01

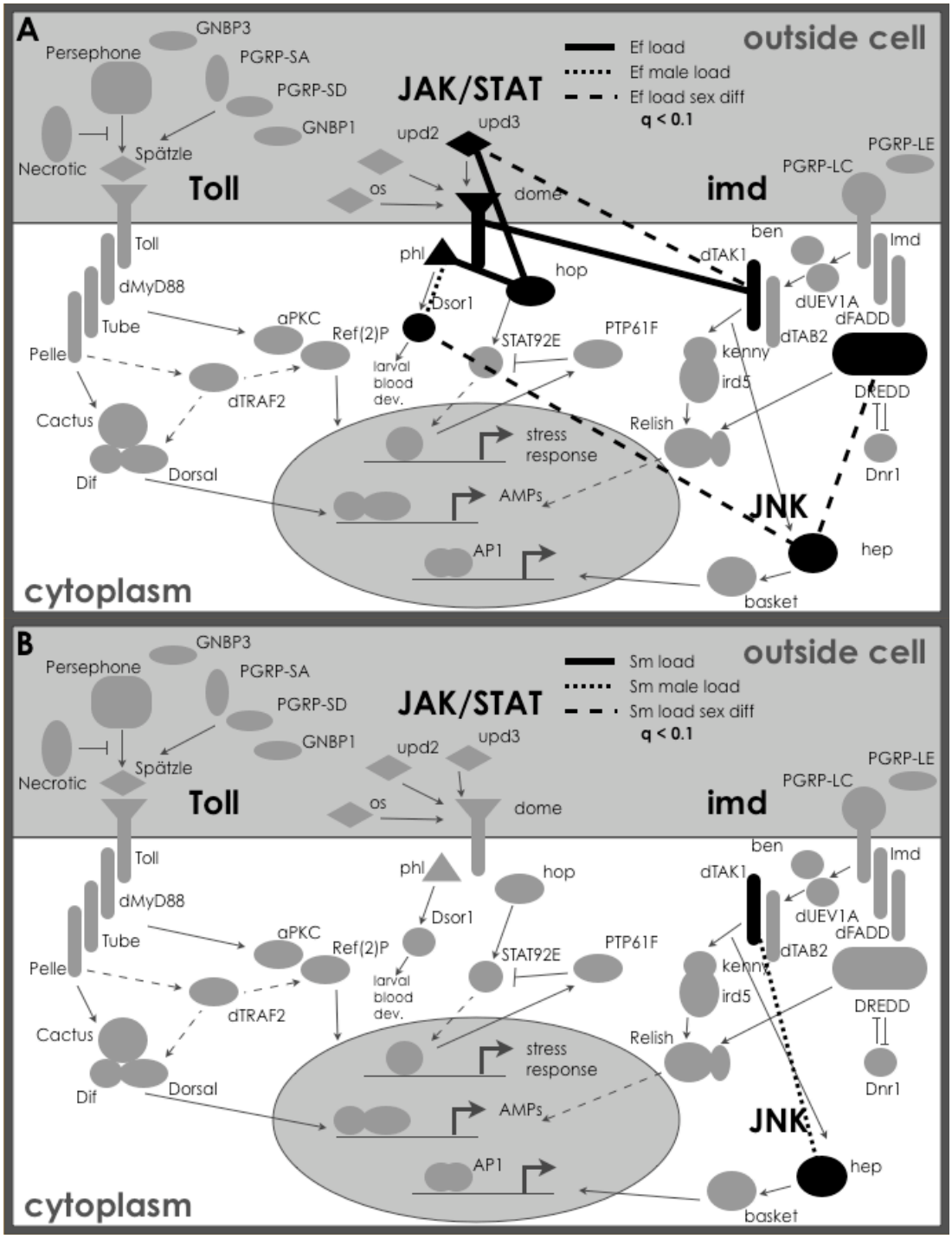


Figure 2.5 Epistatic interactions associating with bacterial load after infection with (A) *E. faecalis* and (B) *S. marcescens*. Lines between a pair of genes correspond to at least one interaction between SNPs in those genes having a significant effect on the load phenotype of the corresponding pattern (ANOVA, $q < 0.1$). Interactions shown associating with male or female load were only those that did not also associate with load in both sexes combined. (Ef = *E. faecalis*; Sm = *S. marcescens*)

bacterial load after infection. Most genes – 11 out of 12 tested in these pathways – contain SNPs involved in interactions associating with one or more of the load phenotypes at the $q < 0.1$ level. Additional interactions were found to be significant at this threshold involving genes outside these pathways (thus not depicted on these diagrams), including *lozenge*, *Pvf1*, *Ser7*, and *Tsf1*. One of the genes, *upd2*, also had an interaction between a pair of SNPs within the same gene that associate with a load phenotype ($q < 0.1$). The interaction terms of the models explain different amounts of phenotypic variance, even among interactions showing significant effects (at the $q < 0.1$ level). Some interactions account for less than 0.1% of the variance, while others explain up to 20.6% (*Pvf1* × *Traf2* interaction term with female load after *S. marcescens* infection). This amount of variance is substantially higher than that explained by each SNP individually; the sum of the percent variance explained by the two included SNPs is less than 0.1% in this instance.

Genetic Variation in Immune Gene Induction: To evaluate the effects of X-linked genetic variation on immune gene induction, a subset of 16 lines was selected from the collection of X-extraction lines, and males and females of these lines were assayed for expression of immune-related genes before and after infection with *E. faecalis*. We used the differences in these levels to quantify the induction of gene expression in response to infection. The genes examined include those encoding the antimicrobial peptides Defensin (Def), DiptericinA (DptA) and Metchnikowin (Mtk), as well as Peptidoglycan Recognition Protein-SA (PGRP-SA) and Transferrin1 (Tsf1), involved in iron transport. The subset of lines was selected from the tails of the distribution of mean sex differences of *E. faecalis* loads, allowing for tests of associations of immune gene induction with sex differences in load.

The 88 immune-related SNPs were tested for association with induction phenotypes for each of these genes, in males, females, and both sexes combined, as

well as in phenotypic associations with sex by SNP interactions. Table 2.4 lists the SNPs that showed association (with FDR < 10%) with one or more of the induction phenotypes. Only 10 of the 88 SNPs appear with significant associations, yet these represent variation in seven separate genes (six of which also have SNPs or haplotypes associating with load phenotypes). While some SNPs show associations with more than one phenotype, most show isolated effects - five out of the nine associate with only one of the induction phenotypes. Most of these associations appear with induction phenotypes in only one sex, yet only one SNP associates significantly with sex difference in induction. While most of these associations explain less than 8% of the variance observed (some less than 1%), the *Ser7* exonic SNP appears to explain more than 14% of the observed variance in *Mtk* induction in females (Table 2.4).

Correlations Between Induction and Load Phenotypes: In addition to testing associations between genetic variation and immune gene phenotypes, we also tested whether any of the variation observed in induction of immune genes correlated with variation in bacterial load after infection with *E. faecalis*. Here, we tested the ability of models incorporating expression levels (before and after infection, as well as levels of induction) to explain levels of bacterial load in these lines of flies. One putative association was found, where the induction of *Tsf1* correlates negatively with bacterial load after *E. faecalis* infection ($P = 0.008$, based on permuted null distribution) in males. Figure 2.6 displays bacterial load line means plotted against induction line means of *Tsf1* (normalized by *RpL32* expression) in both males and females after infection with *E. faecalis*. Increased induction levels of *Tsf1* associate with lower levels of bacterial load after infection in males, while female values show no significant correlation between these traits.

Table 2.4 SNPs Associating with Immune Gene Induction Phenotypes

				Female Induction					Male Induction				
Functional Class	**Gene**	**Location**	**Change**	**Def**	**DptA**	**Mtk**	**PGRP-SA**	**Tsf1**	**Def**	**DptA**	**Mtk**	**PGRP-SA**	**Tsf1**
Signal Transduction	*pole hole*	exon		0.997	0.820	0.506	0.230	0.170	0.551	0.518	0.954	0.976	<0.001*** 3.0
	pole hole	exon		0.997	0.820	0.887	0.230	0.170	0.368	0.197	0.954	0.144	0.056* 1.8
	TRAF2	intergenic		0.023* 6.7	0.870	0.813	0.161	0.247	0.551	0.922	NA	0.976	0.806
	TRAF3	exon		NA	<0.001*** 7.1	NA	0.228	NA	0.382	<0.001*** 4.8	0.954	NA	NA
Secreted	*upd2*	exon		<0.001*** 0.3	0.820	1.000	0.248	0.170	0.551	0.943	0.681	0.976	0.042* 7.9
Iron Metabolism	*Tsf1*	exon		<0.001*** 4.5	NA	0.017* 1.6	0.228	NA	NA	NA	NA	NA	NA
Hematopoesis	*Pvf1*	exon	**A->T**	NA	0.381	NA	NA	NA	0.551	0.518	NA	NA	NA
	Pvf1	exon		0.997	0.820	0.992	0.161	0.075* 1.7	0.551	0.546	NA	0.976	0.806
Serine Protease	*Ser7*	intergenic		<0.001*** 4.6	NA	<0.001*** 1.2	0.228	NA	NA	NA	NA	NA	NA
	Ser7	exon		0.261	0.820	0.046* 14.3	0.272	0.075* 4.5	0.551	0.197	NA	0.976	0.806

				Combined Induction					Effect of Sex by SNP Interaction on Induction				
Functional Class	**Gene**	**Location**	**Change**	**Def**	**DptA**	**Mtk**	**PGRP-SA**	**Tsf1**	**Def**	**DptA**	**Mtk**	**PGRP-SA**	**Tsf1**
Signal Transduction	*pole hole*	exon		0.582	0.682	0.425	0.481	0.277	0.923	0.620	NA	0.759	0.539
	pole hole	exon		0.599	0.645	0.620	0.412	0.387	0.906	0.318	0.760	0.694	0.592
	TRAF2	intergenic		0.594	0.874	0.425	0.549	0.912	0.788	0.832	NA	0.657	0.763
	TRAF3	exon		0.448	<0.001*** 5.9	0.620	0.515	NA	NA	NA	NA	NA	NA
Secreted	*upd2*	exon		0.149	0.874	0.620	0.549	0.387	0.788	0.832	0.760	0.657	0.488
Iron Metabolism	*Tsf1*	exon		0.119	0.899	<0.001*** 1.7	0.515	0.273	NA	NA	NA	NA	NA
Hematopoesis	*Pvf1*	exon	**A->T**	0.849	0.450	NA	0.515	0.657	NA	0.620	NA	<0.001*** 0.3	NA
	Pvf1	exon		0.691	0.450	0.620	0.515	0.912	0.906	0.832	NA	0.657	0.539
Serine Protease	*Ser7*	intergenic		<0.001*** 0.2	0.902	<0.001*** 0.5	0.515	<0.001*** 0.0	NA	NA	NA	NA	NA
	Ser7	exon		0.590	0.450	0.104	0.549	0.908	NA	0.741	NA	0.750	0.539

Q-values shown for SNPs that associate with at least one induction phenotype at an FDR ≤ 10% (q ≤ 0.1); 79 of the 88 SNPs tested showed no association with any induction phenotype. Percent of phenotypic variance explained by association shown for each test with q < 0.1 (below q-value). Change = amino acid change associated with SNP, where applicable. *q < 0.1, *** q < 0.001

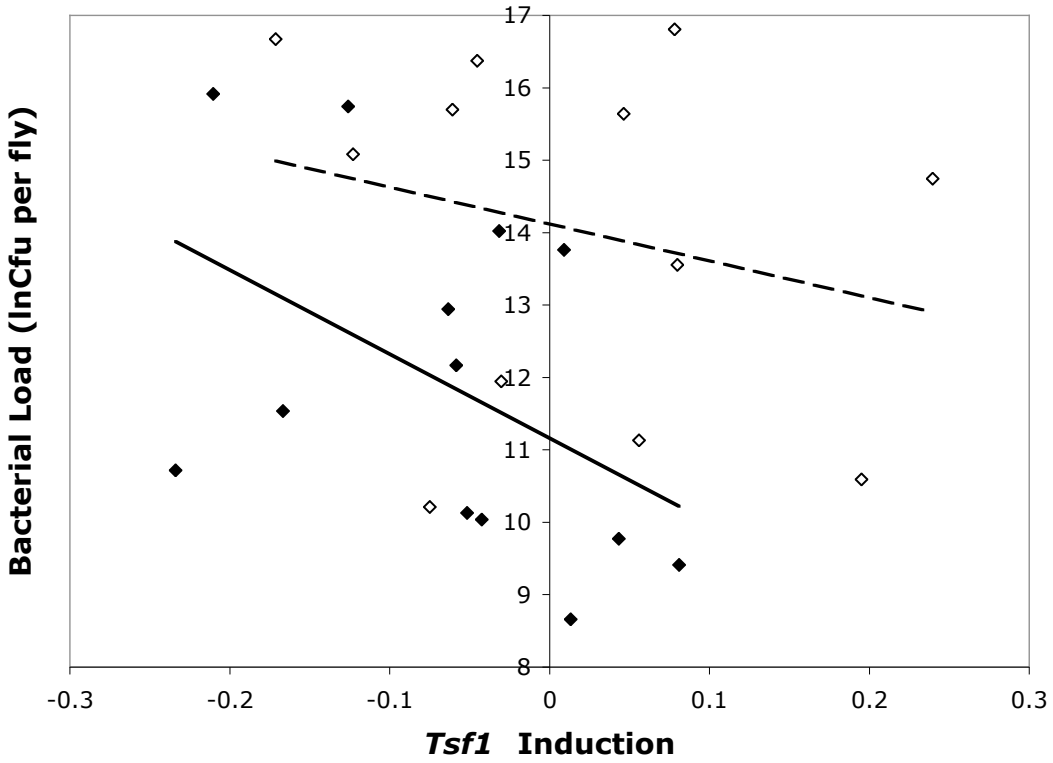


Figure 2.6 Line means for bacterial load vs. *Tsf1* induction levels (infected minus uninfected expression, normalized by *RpL32* expression) after infection with *E. faecalis* in males (closed points, solid regression line) and females (open points, dashed regression line).

Discussion

To examine the effects of genomic location on genotypic variation and phenotype, we measured associations between polymorphisms in X-linked immune genes and response to bacterial infection in lines of *D. melanogaster*. These lines, bearing naturally varying X chromosomes in a co-isogenic autosomal background, were genotyped for SNPs in 20 immune genes. These X-linked immune genes include members of numerous immune-related pathways. The Toll and imd pathways, key to the humoral antimicrobial response in Drosophila, have representatives on the X chromosome, and the JAK/STAT pathway, also involved in response to bacterial infection, has a significant excess of its genes on the X (χ^2, d.f.=1, $P = 7.7 \times 10^{-5}$). Interestingly, while the X-linked genes from these pathways include those with roles in recognition and signaling, there are no antimicrobial peptide genes yet identified on the X chromosome. The absence of antimicrobial peptide genes on the X chromosome is highly significant (Fisher's Exact Test, $P = 0.0036$), and the cause for this remains a puzzle.

Other than the genic content of the X compared to the autosomes, this chromosome also provides a unique environment that may allow different levels and types of genetic variation to exist compared to that on the autosomes. Furthermore, since genes on the X chromosome spend 1/3 of their time in hemizygous males, they are exposed to different selective pressures; hemizygosity may expose recessive alleles, purging deleterious genotypes and fixing beneficial ones. This is expected to result in lower levels of variation on the X chromosome relative to autosomes with the exception of alleles showing antagonistic phenotypes either between sexes or in different environments or genetic backgrounds (CHARLESWORTH *et al.* 1987).

For immune-related genes on the X chromosome, we expect that variation may be maintained in the population more readily if different alleles provide beneficial

effects in diverse environments, such as with different bacterial infections, or in distinct genetic backgrounds, including in males *vs*. females. The results found in this study agree with this expectation. Genetic variation is observed in X-linked immune genes, frequently associating with phenotypic variation in immune response. Many of these associations, though, appear with one bacterial infection and not the other, or act in a sex-specific or sexually antagonistic manner. Alleles such as these, associating with phenotypic variation in a condition-specific manner, presumably would not be selected for or against as rapidly as those with universally beneficial or deleterious effects, even on the X chromosome. A few SNPs tested here do show more general associations with the immune phenotypes examined; presumably alleles that appear relatively detrimental in tests observed here could have been maintained in the population because of beneficial effects in other circumstances (or for other phenotypes).

Previous investigations (LAZZARO *et al.* 2006; LAZZARO *et al.* 2004) have involved similar tests of association between genotypic variation in immune genes on the second and third chromosomes of *D. melanogaster* and differences in immune response phenotypes. It is difficult to make direct comparisons between these studies, involving fly lines from separate populations, different bacterial infections, and distinct experimental setups, including different levels of replication. We do find, though, that similar to those studies, variation in numerous genes throughout immune pathways associate significantly with phenotypic variation. Interestingly, genetic variation on the second chromosome can explain 47.2% of the total variance in bacterial load (after infection with *S. marcescens*, (LAZZARO *et al.* 2004)), and variation on the third chromosome can explain 22.1% of the total observed variance in bacterial load (after infection with *P. rettgeri*, (SACKTON *et al., submitted*)), yet X-linked genetic variation in these lines explains only 15.5% of the total variance in

bacterial load (after infection with *S. marcescens*). This suggests a lower level of naturally occurring variation in X-linked immune genes and/or less influence of that variation on immune phenotypes than is observed with autosomal genes. Additionally, this could be due to the fact that the X chromosome, as a shorter portion of the genome compared to the autosomes, may simply contain fewer loci affecting the observed immune phenotypes. Individual polymorphisms within immune genes on the second chromosome, though, appear to explain a larger proportion of the phenotypic variance than was observed with the SNPs here; numerous autosomal variations explained more than 5% of the phenotypic variance, whereas the most significant explained up to 22.7% (LAZZARO *et al*. 2006, LAZZARO *et al*. 2004). Thus, variation in X-linked immune genes could have a lesser influence on phenotypic variance than polymorphism in autosomal genes, relative to environmental and experimental factors influencing differences in bacterial load.

The associations found between autosomal genes and immune phenotypes were strongly biased with respect to the functional class of immunity genes. There was a preponderance of associations between bacterial load and SNPs in recognition molecules, and a deficit of associations with SNPs in antimicrobial peptides (LAZZARO *et al.* 2004). The X chromosome had a markedly different distribution of functional classes of immune genes, including an absence of any antimicrobial peptides, and this may contribute to the observation that there was no departure from random representation of recognition and signaling functional classes among X-linked genes that associated with bacterial defense (χ^2, d.f. = 1, $P = 0.327$). Furthermore, since variation in autosomal antimicrobial peptide genes appears to lack the phenotypic associations that variation in other autosomal immune genes contain, it seems unlikely that genic makeup alone would lead to different patterns of association between the X-linked and the autosomal immune genes.

While polymorphisms in X-linked genes appear to act mostly in sex-specific or sexually antagonistic associations with phenotypic variation, associations involving autosomal variation were much more likely to be sex-independent. An investigation of associations with variation in immune genes on the second chromosomes uncovered no significant sex or sex × line effects on load (LAZZARO *et al.* 2004). Similarly, SNPs on the third chromosome associating with immune phenotypes show only marginal effects of sex × SNP interaction (SACKTON *et al*., *submitted*). The inflated magnitude of sexual dimorphism of X-linked immunity genes over autosomal genes is consistent with several mechanisms that might produce different regulatory responses of X-linked genes, including unique patterns of sex-biased expression (PARISI *et al.* 2003) or imprecision of dosage compensation.

In addition to dimorphic effects on correlations between genotype and load, we also observe sex-specific associations between induction levels and load in these lines. This suggests that activation of parts of the immune response may be regulated in sex-specific manners. This may reflect differences between the sexes in immune response needs and/or priorities. Since Drosophila males and females may have different fitness impacts of immune system activation (MCKEAN and NUNNEY 2001; MCKEAN and NUNNEY 2005), it is reasonable to expect some genotypes to have different phenotypic effects between the sexes.

As well as the sexual dimorphism that appears in the associations between X-linked genetic variation and immune response phenotypes, distinct responses to different species of bacterial infections were also observed here. Four out of the 19 SNPs (21.1%) that associate with bacterial load phenotypes in these lines (in both sexes combined, with $P < 0.05$), though, show associations with variation in response to both *E. faecalis* and *S. marcescens*, while only one out of 36 (2.8%) of all of the autosomal SNPs associating with load differences after infections with one of these

two bacteria are commonly found between the two. This excess of overlapping associations among X-linked polymorphisms over those on the autosomes is marginally significant (Fisher's Exact Test, $P = 0.048$), indicating a higher level of generality in the X-linked associations with response to different bacterial infections. Most of the variation tested in these lines (both X-linked and autosomal), though, appears to have the same effect across infections; flies bearing an allele that associates with lower load after infection with one bacterium tend to have lower load after the infection with the other bacterium as well. Thus, while the variation among X-linked and autosomal immune genes may vary in generality of response to different bacteria, there is not much evidence for antagonistic variation between bacterial infections. The widespread presence of sex differences in associations in this study underscores the complexity of the association between immune response and polymorphisms in X-linked immune genes. These effects impacting genotype-phenotype correlations appear to be more striking with X-linked variation than with that on the autosomes; this is not unexpected, though, given the genomic environments of genes on these respective chromosomes. If the X-linked variation existing in natural populations includes alleles detrimental in one sex but not the other, these alleles are less likely to be selected against and may remain in the population in spite of negative phenotypic effects. Thus, even though we may expect genotypic variation associating with phenotypic effects to be relatively uncommon on the X chromosome, phenotypic differences observed here do correlate with polymorphisms in these lines. The complex patterns of association seen, however, show that these segregating polymorphisms bear characteristics consistent with predicted effects of natural selection on X-linked variation.

REFERENCES

ANDOLFATTO, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol Biol Evol **18:** 279-290.

ARBOUZOVA, N. I., and M. P. ZEIDLER, 2006 JAK/STAT signalling in Drosophila: insights into conserved regulatory and cellular functions. Development **133:** 2605-2616.

BARRETT, J. C., B. FRY, J. MALLER and M. J. DALY, 2005 Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics **21:** 263-265.

BRENNAN, C. A., and K. V. ANDERSON, 2004 Drosophila: the genetics of innate immune recognition and response. Annu Rev Immunol **22:** 457-483.

CARBONE, M. A., K. W. JORDAN, R. F. LYMAN, S. T. HARBISON, J. LEIPS *et al.*, 2006 Phenotypic variation and natural selection at *catsup*, a pleiotropic quantitative trait gene in Drosophila. Curr Biol **16:** 912-919.

CHARLESWORTH, B., J. A. COYNE and N. H. BARTON, 1987 The Relative Rates of Evolution of Sex Chromosomes and Autosomes. The American Naturalist **130:** 113-146.

CHEVERUD, J. M., and E. J. ROUTMAN, 1995 Epistasis and its contribution to genetic variance components. Genetics **139:** 1455-1461.

CLARK, A. G., and L. WANG, 1997 Molecular population genetics of Drosophila immune system genes. Genetics **147:** 713-724.

COCKERHAM, C. C., 1954 An Extension of the Concept of Partitioning Hereditary Variance for Analysis of Covariances among Relatives When Epistasis Is Present. Genetics **39:** 859-882.

DATE, A., Y. SATTA, N. TAKAHATA and S. I. CHIGUSA, 1998 Evolutionary history and mechanism of the Drosophila *Cecropin* gene family. Immunogenetics **47:** 417-429.

DE GREGORIO, E., P. T. SPELLMAN, G. M. RUBEN and B. LEMAITRE, 2001 Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. Proc Natl Acad Sci U S A **98:** 12590-12595.

FERRANDON, D., J. L. IMLER, C. HETRU and J. A. HOFFMANN, 2007 The Drosophila systemic immune response: sensing and signalling during bacterial and fungal infections. Nat Rev Immunol **7:** 862-874.

FOLEY, E., and P. H. O'FARRELL, 2004 Functional dissection of an innate immune response by a genome-wide RNAi screen. PLoS Biol **2:** E203.

GIBSON, J. R., A. K. CHIPPINDALE and W. R. RICE, 2002 The X chromosome is a hot spot for sexually antagonistic fitness variation. Proc Biol Sci **269:** 499-505.

IRVING, P., L. TROXLER and C. HETRU, 2004 Is innate enough? The innate immune response in Drosophila. C R Biol **327:** 557-570.

JIGGINS, F. M., and G. D. HURST, 2003 The evolution of parasite recognition genes in the innate immune system: purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. J Mol Evol **57:** 598-605.

KIMBRELL, D. A., and B. BEUTLER, 2001 The evolution and genetics of innate immunity. Nat Rev Genet **2:** 256-267.

LAZZARO, B. P., 2005 Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans*. Genetics **169:** 2023-2034.

LAZZARO, B. P., and A. G. CLARK, 2003 Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. Mol Biol Evol **20:** 914-923.

LAZZARO, B. P., T. B. SACKTON and A. G. CLARK, 2006 Genetic variation in *Drosophila melanogaster* resistance to infection: a comparison across bacteria. Genetics **174:** 1539-1554.

LAZZARO, B. P., B. K. SCEURMAN and A. G. CLARK, 2004 Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. Science **303:** 1873-1876.

LECLERC, V., and J. M. REICHHART, 2004 The immune response of *Drosophila melanogaster*. Immunol Rev **198:** 59-71.

LONG, A. D., R. F. LYMAN, C. H. LANGLEY and T. F. MACKAY, 1998 Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. Genetics **149:** 999-1017.

MCKEAN, K. A., and L. NUNNEY, 2001 Increased sexual activity reduces male immune function in *Drosophila melanogaster*. Proc Natl Acad Sci U S A **98:** 7904-7909.

MCKEAN, K. A., and L. NUNNEY, 2005 Bateman's principle and immunity: phenotypically plastic reproductive strategies predict changes in immunological sex differences. Evolution **59:** 1510-1517.

MEDZHITOV, R., and C. A. JANEWAY, JR., 1997 Innate immunity: impact on the adaptive immune response. Curr Opin Immunol **9:** 4-9.

PARISI, M., R. NUTTALL, D. NAIMAN, G. BOUFFARD, J. MALLEY *et al.*, 2003 Paucity of genes on the Drosophila X chromosome showing male-biased expression. Science **299:** 697-700.

RAMOS-ONSINS, S., and M. AGUADE, 1998 Molecular evolution of the *Cecropin* multigene family in Drosophila. functional genes vs. pseudogenes. Genetics **150:** 157-171.

RICE, W. R., 1984 Sex Chromosomes and the Evolution of Sexual Dimorphism. Evolution **38:** 735-742.

ROZAS, J., and R. ROZAS, 1995 DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. Comput Appl Biosci **11:** 621-625.

SACKTON, T. B., B. P. LAZZARO, T. A. SCHLENKE, J. D. EVANS, D. HULTMARK *et al.*, 2007 Dynamic evolution of the innate immune system in Drosophila. Nat Genet **39:** 1461-1468.

SCHEET, P., and M. STEPHENS, 2006 A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet **78:** 629-644.

SCHLENKE, T. A., and D. J. BEGUN, 2003 Natural selection drives Drosophila immune system evolution. Genetics **164:** 1471-1480.

SCHMID-HEMPEL, P., 2003 Variation in immune defence as a question of evolutionary ecology. Proc Biol Sci **270:** 357-366.

SINGH, N. D., A. M. LARRACUENTE and A. G. CLARK, 2008 Contrasting the efficacy of selection on the X and autosomes in Drosophila. Mol Biol Evol **25:** 454-467.

STOREY, J. D., 2002 A direct approach to false discovery rates. J. R. Statist. Soc. B **64:** 479-498.

STRONACH, B., and N. PERRIMON, 2002 Activation of the JNK pathway during dorsal closure in Drosophila requires the mixed lineage kinase, slipper. Genes Dev **16:** 377-387.

VILELLA, A. J., A. BLANCO-GARCIA, S. HUTTER and J. ROZAS, 2005 VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. Bioinformatics **21:** 2791-2793.

WASSARMAN, D. A., M. THERRIEN and G. M. RUBIN, 1995 The Ras signaling pathway in Drosophila. Curr Opin Genet Dev **5:** 44-50.

CHAPTER 3

FUNCTIONAL REGULATORY DIVERGENCE OF THE INNATE IMMUNE SYSTEM IN INTERSPECIFIC DROSOPHILA HYBRIDS

Abstract

In order to investigate divergence of immune regulation among Drosophila species, we have engaged in a detailed study of innate immune function in F1 hybrids of *D. melanogaster* and *D. simulans*. If pathways have diverged between the species such that incompatibilities have arisen between interacting components of the immune network, we expect the hybrids to display dysregulation of immune genes. We have quantified gene induction in hybrid and parental flies in response to bacterial infection. These results show that while the hybrid flies do not suffer widespread immune breakdown, they show significantly different regulation of many immune genes relative to the parents. We examine this divergence in terms of additivity and expression differences among genes, observing distinct patterns of dysregulation among functional groups within the pathways of the innate immune system. The functional groups most sensitive to misexpression in the hybrids are the downstream components of the network, indicative of some propagation of dysregulation throughout the immune pathways. Interestingly, this dysregulation does not appear to associate with phenotypic differences in bacterial load post-infection in hybrids, possibly highlighting some robustness of function of innate immune response to perturbations like hybridization.

Introduction

Innate immunity has garnered interest in numerous areas of biological research due to its ubiquity throughout the animal kingdom; in addition to serving as the only

response to infection in invertebrates, the innate immune system provides an initial, generalized attack against invading microbes and activates the adaptive immune response in humans and other vertebrates (MEDZHITOV and JANEWAY 1997). The deep evolutionary conservation of these pathways across taxa underscores the importance of maintaining their components and functions for proper immune response. Nevertheless, divergence in the innate immune response is evident – even between closely related species – in the form of nucleotide or amino acid sequence differences among orthologs, as well as at the level of which genes and gene families comprise the immune pathways (DATE *et al*. 1998; SACKTON *et al*. 2007). This may reflect distinct pathogen environments, driving diverse selection pressures and different roles of immune response in the context of life history traits among species (SCHMID-HEMPEL 2003). If components of the innate immune system have diverged between two closely related Drosophila species, we expect to see disruption of immune response regulation in hybrids between those species. In this study, we quantify levels of dysregulation in the hybrid immune response to pinpoint regulatory divergence between the parental species.

Drosophila has become an effective model system for the investigation of the evolution of innate immunity due to the well-studied genetics of insect response to infection and the comparative resources available. While immune genes as a group may undergo greater levels of positive selection than non-immune genes in flies (SCHLENKE and BEGUN 2003), it has also become evident that when divided among more specific functions, different subgroups of immune genes show distinct patterns of evolution. For example, antimicrobial peptides (AMPs), despite facing direct contact with invading microbial cells, have shown little amino acid divergence among Drosophila species (CLARK and WANG 1997; DATE *et al*. 1998; LAZZARO and CLARK 2003; RAMOS-ONSINS and AGUADE 1998). On the other hand, some immune-related

recognition proteins (scavenger receptors) show evidence of rapid evolution (LAZZARO 2005). More recently, large-scale examination of patterns of selection acting upon each of the immune genes in the genomes of multiple Drosophila species has become possible. (SACKTON *et al*. 2007) used the newly sequenced genomes of six species in the melanogaster subgroup to make sweeping comparisons of immune genes. They found further evidence of distinct patterns of selection among different functional classes of immune genes. In addition to the wealth of sequence data available for the genomes of these Drosophila species, the use of genome-wide expression arrays has allowed for investigation into the details of the regulation of immune response in flies. Numerous studies using these arrays before and after microbial infection (APIDIANAKIS *et al*. 2005; DE GREGORIO *et al*. 2001; IRVING *et al*. 2001) have clarified the dynamics of immune response, solidifying existing models of innate immunity as well as identifying new genes and pathways that are regulated in response to infection.

Throughout an organism, there are numerous systems whose regulatory components may coevolve (DOVER 1992; SHAW *et al*. 2002). While maintaining proper regulation within a species, if the interactions have diverged separately in two lineages, these may yield incompatibilities in the context of interspecific hybrids (LANDRY *et al*. 2007; TRUE and HAAG 2001). Divergence between closely related species has been inferred from irregular development or enzyme expression patterns in interspecific hybrids (DICKINSON *et al*. 1984; PARKER *et al*. 1985; WHITT *et al*. 1973; WHITT *et al*. 1977), and more recently, regulatory divergence between species has been estimated through the quantification of genome-wide expression levels in F1 hybrids using microarrays (AUGER *et al*. 2005; MICHALAK and NOOR 2003; MOEHRING *et al*. 2007; RANZ *et al*. 2004). These studies have revealed numerous instances of non-additivity of expression in hybrid individuals relative to parental

phenotypes, indicative of specifically evolved regulatory mechanisms. Furthermore, in complex regulatory networks, one would expect hybrids to be particularly prone to dysregulation; if one or more portions of a pathway have diverged between parental species such that they result in incompatibilities in the hybrids, these may propagate throughout the network, manifesting in large-scale disruptions of regulatory phenotypes.

Since the innate immune pathways in Drosophila contain many interacting components and regulatory elements, it is likely that these may have diverged between species and that interspecific hybrids may bear phenotypes reflecting this divergence. To investigate divergence of the regulatory elements of the innate immune system between Drosophila species, we have quantified dysregulation of the immune response in interspecific hybrids by assaying transcription abundance in *D. melanogaster*, *D. simulans*, and their F1 hybrids before and after infection. If genes throughout the innate immune pathways have diverged between these two species, we expect the hybrids to display non-additive levels of expression. Additionally, if different portions of these networks bear different levels of interspecific divergence, the patterns of dysregulation throughout the hybrid immune response should reflect those differences.

Materials and Methods

Fly Lines and Crosses: Inbred stocks of *Drosophila melanogaster zygotic hybrid rescue (zhr)* strain (provided by A. Orr) and a Tsimbazaza strain of *D. simulans* (provided by H. Hollocher) were used to construct hybrid crosses. Lines of each were maintained in lab cultures, and from these we collected *D. melanogaster* virgin females and *D. simulans* males. Interspecific crosses were set up with approximately 10 *D. melanogaster* females and 10 *D. simulans* males per vial. Intraspecific crosses were also set up simultaneously, with about 10 females and 10 males apiece, to

produce *D. melanogaster* and *D. simulans* offspring under similar conditions and at similar ages as the F1 hybrid flies. F1 hybrid female flies (male hybrids from this cross are not viable), along with *D. melanogaster* and *D. simulans* female flies, were collected after eclosion.

Bacterial Cultures and Infections: To assay response to bacterial infection in the flies, we infected them with Gram-negative *Serratia marcescens*. This bacterium, chosen based on its previous use for immune studies in *D. melanogaster* (LAZZARO *et al*. 2004), was derived from ATCC strain 13880. Bacterial cultures for infections were grown overnight from freezer stocks, to a concentration of $OD_{600} \approx 1.0$. Female offspring from *D. melanogaster*, *D. simulans* and the *D. mel* × *D. sim* hybrid cross were infected at approximately three to seven days after eclosion. Flies were infected by pricking their thoraces with 0.1-mm tungsten needles (Fine Science Tools, Foster City, CA) dipped in bacterial culture.

Transcript Quantification Using BeadChip Arrays: To estimate expression differences before and after bacterial infection in hybrid flies and those from parental strains, we used custom BeadChip Arrays (Illumina Inc., San Diego, CA) to quantify transcript abundance in the samples. These were designed to include probes for 171 immune-related genes, along with 542 genes representing controls or pathways investigated in other experiments (SACKTON *et al*., *submitted*). See Table B.1 for a full list of genes included on the BeadChips, including classification of immune-related genes by functional group. Flies from each strain were snap-frozen in liquid nitrogen at six and at 12 hours after infection, in three replicate pools of approximately 12-15 flies each. Uninfected flies were also frozen immediately after infection to measure baseline expression levels. For each sample, we isolated mRNA using a Trizol:chloroform extraction, and then we synthesized cDNA and hybridized it to the BeadChips using the BeadChip protocol. BeadChips were scanned, and the resulting

signal values were normalized across arrays and across chips using qspline in the beadarray R package.

Bacterial Load Quantification: Following previous studies (LAZZARO *et al*. 2004), bacterial clearing ability was estimated by quantifying bacterial load in infected flies. Approximately 12 or 25 hours after infection, *D. melanogaster*, *D. simulans*, and F1 hybrid flies were homogenized, three at a time, in 500 μl of LB broth. These samples (n = 4-9 for each line and time point) were then plated onto agar plates using a spiral plater (Spiral Biotech, Bethesda, MD). Plates were kept overnight to allow colonies to grow enough to be counted by a colony counter to infer bacterial concentration inside each homogenate sample. Plates were visually inspected to ensure that colonies counted showed size and morphology expected.

Statistical Analysis: To test for expression differences before and after bacterial infection, we used mixed linear models incorporating infection status as a fixed effect, along with other random effects:

$$y_{ijk} = \mu + \text{Infection}_i + \text{Probe}_j + \text{Replicate}_k + \varepsilon_{ijkl} \quad (1)$$

Here, y is equal to transcript level, and Infection (i = 1,2) represents the infection status of the flies (either uninfected or infected), included as a fixed effect. Probe (j = 1,2), representing the two separate probes for each gene on the array, and Replicate (k = 1...3) were each included as random effects. Transcript abundance is estimated by $\log_2$(Signal), where Signal is the normalized measurement from the arrays, log-transformed to achieve a more normal distribution. These tests were performed separately for each individual gene, at both six and 12 hours after infection, and for *D. melanogaster*, *D. simulans* and F1 hybrid flies. In order to assess significance of the results of these tests, we compared the coefficients of the infection effect to a null

distribution comprised of coefficients calculated from tests of the same model with expression data permuted 1,000 times relative to the genotype and infection status for each gene.

To evaluate differences in expression levels among these groups of flies, we used similar mixed models to test for significant effect of group or "species" on expression. The models were set up as follows:

$$y_{ijk} = \mu + \text{Species}_i + \text{Probe}_j + \text{Replicate}_k + \varepsilon_{ijkl} \quad (2)$$

In this case, y is again equal to transcript level, and Species (i = 1...3) is a fixed effect including the three groups of flies, *D. melanogaster*, *D. simulans* and F1 hybrids. Probe and Replicate are both included as random effects as in Equation 1. Here, we tested differences among flies for each gene on the chip, examining each treatment type and time point separately. As above, significance for each test was determined by comparison to a null distribution of coefficients calculated using data permuted 1,000 times.

In addition to expression differences among flies, we also quantified induction differences, where induction represents the change in expression before and after infection. To achieve this, we employed similar mixed models, including an interaction term to test the epistatic effect of treatment status (infected vs. uninfected) on species differences in expression:

$$y_{ijklm} = \mu + \text{Species}_i + \text{Treatment}_j + (\text{Species}\times\text{Treatment})_k + \text{Probe}_l + \text{Replicate}_m + \varepsilon_{ijklmn} \quad (3)$$

using the same setup as Equation 2, with the addition of Treatment ($j = 1,2$), representing the infection status of the flies, along with an interaction term between this and Species groups as fixed effects. These interactions were examined separately for six hours after infection and 12 hours after infection, for each gene individually. Once again, we permuted the expression values 1,000 times and collected coefficients for the interaction term against which we could compare the actual results to determine the significance for each test.

For each gene showing differences in expression or induction among the groups of flies, we tested for the presence of non-additive expression (or induction) in the F1 hybrids relative to parental levels. The null hypothesis was that F1 expression levels were equal to midparent expression values, indicative of entirely additive effects. To test the validity of this hypothesis for each gene, we performed tests using models set up like Equation 2, where the only "species" groups included were parental (*D. melanogaster* and *D. simulans* samples combined) and F1 hybrid flies. Significant differences ($P < 0.05$) between parental mean and hybrid expression values allowed us to reject the hypothesis of complete additivity. For these genes, we tested whether hybrid expression showed evidence of dominance or transgressive variation. In this case, the null hypothesis was that F1 expression levels equaled one of the parental expression levels, indicative of dominant effects of that parental allele. We tested for differences among groups of flies as above, but only including expression values for hybrids and one parental species at a time for this set of genes. Significant differences between hybrids and parental values provided evidence for transgressive effects, while no difference here left us unable to reject the hypothesis of dominance of expression in the hybrids.

To compare induction patterns of genes belonging to the Toll and imd pathways in the context of their functional groupings within the pathways, we

quantitatively organized genes by degree of similarity in patterns of expression changes after infection using a simple hierarchical agglomerative clustering algorithm (Cluster 3.0, DE HOON *et al*. 2004). Genes were grouped with hierarchical, centroid linkage clustering using Euclidean distances to determine similarity. In addition, we included an expected gene order based on relative location in the humoral pathways. This informed the clustering such that once nodes of the gene tree were defined, the orientation of each node ordered the genes according to pathway location, wherever possible. When the trees for each group of genes were defined, we used Java TreeView 1.1.3 (SALDANHA 2004) to visualize the gene clusters and associated dendrograms. To quantify clustering of gene expression patterns relative to layout of the genes in the Toll and imd pathways, we calculated the correlation coefficients of gene orders from cluster analysis with those in the immune network. To determine significance of each correlation value, we permuted the clustered gene orders 1,000 times and calculated correlations for each to achieve a null distribution, and then we calculated *P*-values for the actual correlation coefficients based on these distributions.

Results

Expression Changes After Infection: To examine differences in immune response between hybrid and parental flies, we quantified transcript abundance for genes related to innate immunity and other pathways (including those involved in metabolism and reproduction) using custom Illumina BeadChip arrays. Induction (or repression) levels of genes represented on the arrays were estimated by comparing transcript levels before and after infection. Genes showing significant differences (with a nominal $P < 0.05$) in transcript abundance between uninfected and infected flies (at six or 12 hours after infection) were considered to be induced or repressed. Out of all immune-related genes on the array, 14.8% of the tests for all three groups of

flies (76 out of 513) at 6 hours and 8.8% of tests (45 out of 513) at 12 hours had $P < 0.05$, both of which include significantly more than the 5% that would be expected by chance (χ^2, d.f. = 1, $P_{6hr} = 1.99 \times 10^{-24}$, $P_{12hr} = 8.86 \times 10^{-5}$); however, given the fact that many of these genes were selected for their previously observed induction in flies after infection, this is not surprising. Non-immune genes on the array also show some expression changes, yet these are not as substantial, with 8.2% of tests (89 out of 1089) at 6 hours and only 1.6% of tests (17 out of 1089) at 12 hours bearing significantly different expression levels before and after infection with $P < 0.05$. Distributions of *P*-values for tests of expression differences before and after infection are shown in Figure 3.1A for immunity genes and for the genes whose products function in either metabolism or reproduction. Genes in these groups bear significantly different distributions of *P*-values, with substantially more immune genes having induction tests with $P < 0.05$ than either of the other classes of genes ($\chi 2$, d.f. = 2, $P = 4.55 \times 10^{-8}$). Furthermore, even within the immune group, genes corresponding to different functions also show distinct levels of expression changes after infection ($\chi 2$, d.f. = 2, $P = 3.50 \times 10^{-18}$, Figure 3.1B).

Since numerous investigations have quantified expression differences in Drosophila after bacterial infection using whole genome arrays, we compared our results to data from three of these (APIDIANAKIS *et al*. 2005; DE GREGORIO *et al*. 2001; IRVING *et al*. 2001) to evaluate the uniqueness of the genes induced here. Out of the set of genes showing expression changes in our study, only 37.6% (64 out of 170) of the genes with differences at six hours and 54% (34 out of 63) with differences at 12 hours were also found to be induced or repressed after infection in one or more of the other screens (based on the definitions of significant expression differences given in each paper). Similarly, out of the 227 genes on the BeadChips that had been shown to be induced in one of these previous studies, 39.6% (90 genes) had significant

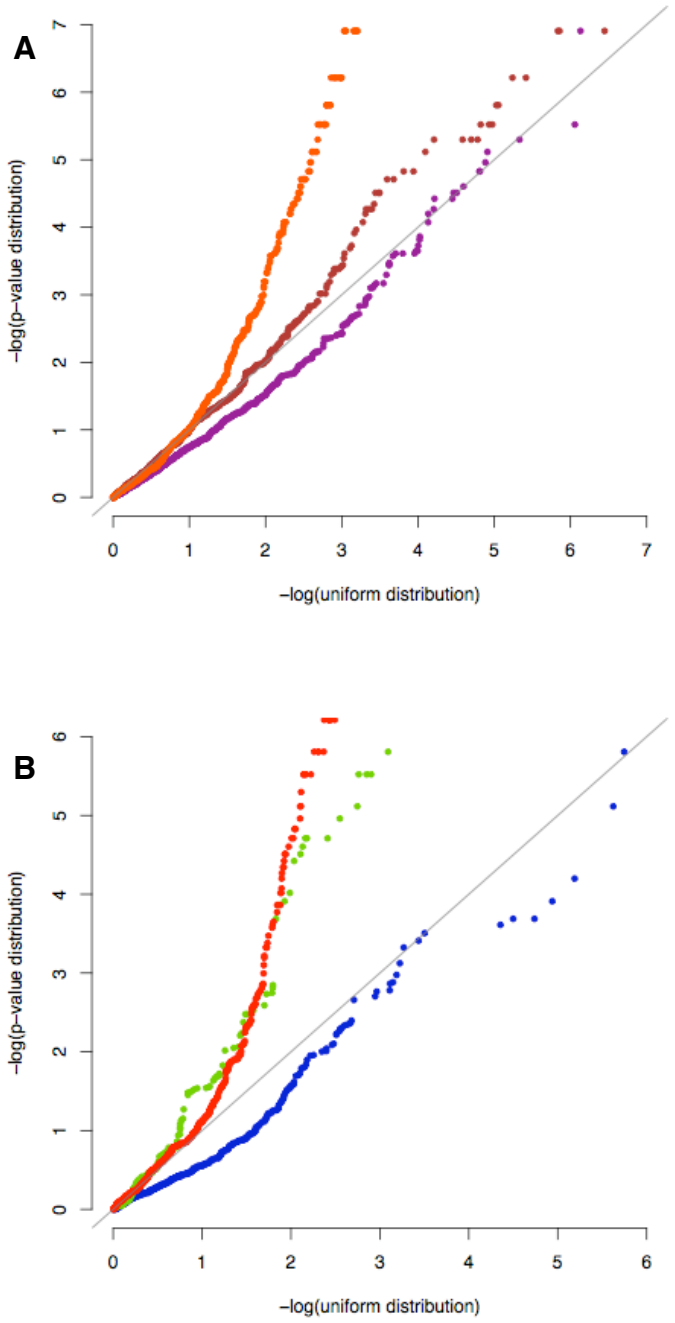


Figure 3.1 Distributions of *P*-values for tests of expression differences before and after infection in groups of genes. Plot A shows immunity (orange), metabolism (brown) and reproduction (purple) genes. Plot B displays functional groups within immunity genes: recognition (green), signaling (blue) and effector (red) genes. Distributions of groups in both A and B vary significantly in the proportion of genes with $P < 0.05$ (χ^2, d.f. = 2, $P = 4.55 \times 10^{-8}$ & $P = 3.50 \times 10^{-18}$, respectively).

differences in expression after infection in this experiment. Causes of these differences may include genetic variation among the lines used and a difference in expression technologies. Most previous investigators have relied on Affymetrix expression microarrays, and there will inevitably be false negatives when contrasting this genome-wide approach with our more focused and sensitive test of immune related genes. There were also differences among previous studies, due to many experimental variables besides the expression platform, including different methods of infection, different timing, and genetic differences among the Drosophila stocks. Furthermore, these studies assayed response to infection with a variety of bacteria; infections with different species of varying levels of pathogenicity in the flies will inevitably have different influences on patterns of expression genome-wide.

Patterns of Non-Additivity in F1 Hybrid Expression: In addition to comparing the number of genes changing expression after infection in the three groups of flies, we also examined levels of expression of genes in F1 hybrids relative to those in *D. melanogaster* and *D. simulans*. To quantify regulatory differences between hybrids and parental species in the expression of immune-related genes, we classified all genes with expression or induction differences among fly groups as showing additive, dominant, or transgressive patterns (Table B.1). We find a substantial number of immune genes showing non-additive expression levels in hybrids – in uninfected flies as well as after infection. Interestingly, not only do levels of non-additivity in immune genes vary across time points, but immune genes in distinct functional roles (recognition, signaling, or effector) also show diverse patterns of additive, dominant, and transgressive expression, as shown in Figure 3.2. Here we observe that expression levels of most of these immune genes – both before and after infection – are non-additive in F1 hybrids, relative to parental levels. On the other hand, induction/repression levels (expression differences before and after infection)

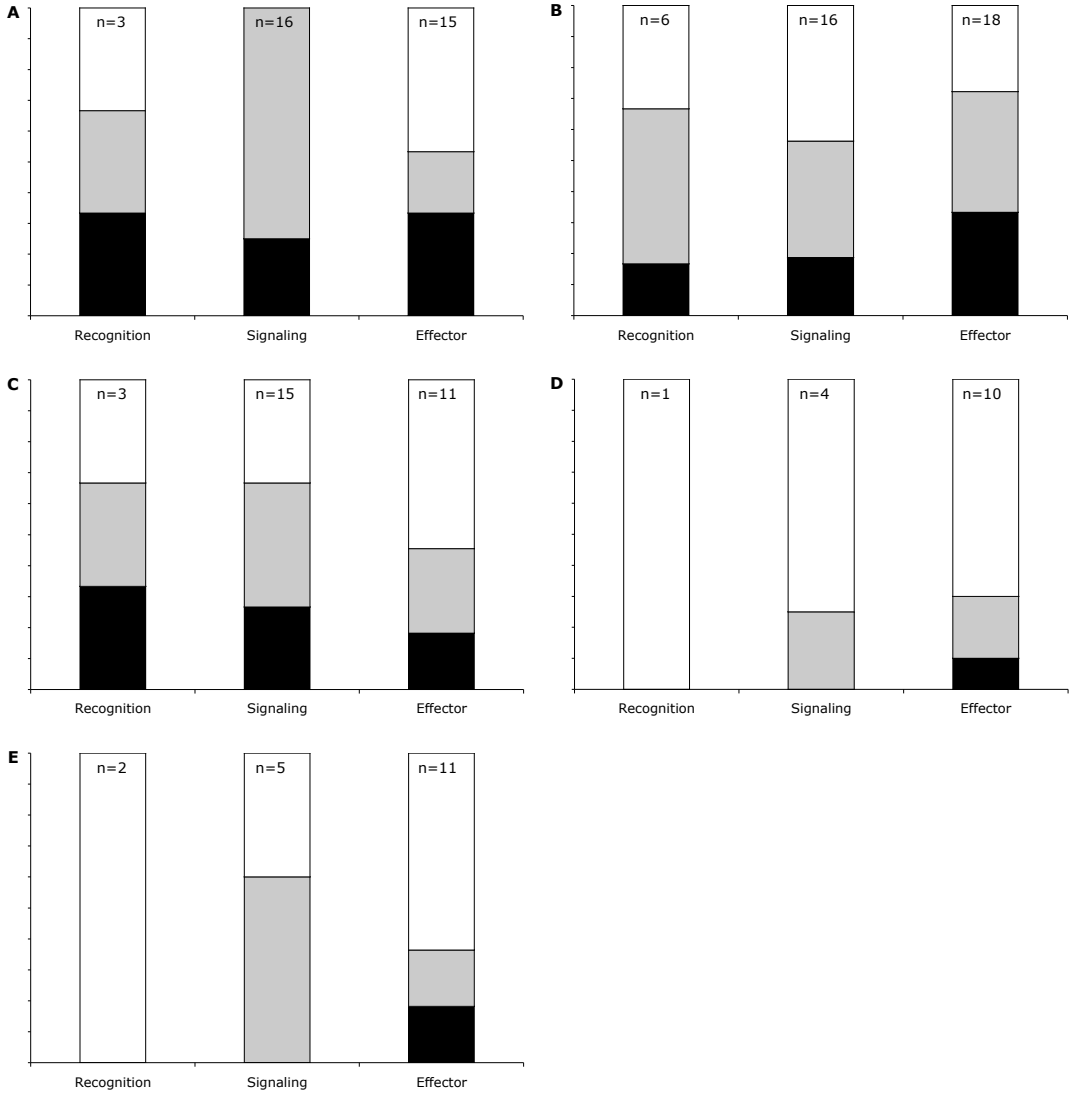


Figure 3.2 Proportion of immune genes differently expressed among *D. melanogaster*, *D. simulans* and F1 hybrid flies in recognition, signaling and effector classes showing patterns of additive (white), dominant (gray) or transgressive (black) expression patterns in the F1 hybrids. A) Expression in uninfected flies. B) Expression in flies 6 hours after infection with *S. marcescens*. C) Expression in flies 12 hours after infection with *S. marcescens*. D) Induction/repression in flies 6 hours after infection with *S. marcescens*. E) Induction/repression in flies 12 hours after infection with *S. marcescens*. Y-axes for plots represent proportion of all genes differently expressed among groups of flies belonging to each pattern of expression; numbers at the top of each column indicate total number of genes in that column.

appear to be much more conserved in these flies; fewer genes show differences in levels of expression change among groups of flies, and most genes that are differently induced/repressed show additive effects in hybrids. In fact, only genes with effector functions show evidence for transgressive effects of induction or repression; other genes that differ among species groups only display additive or dominant effects.

Expression Levels Among Different Functional Classes of Immune Genes in F1 Hybrids: To further examine the differences between hybrid and parental mean expression levels indicated by the non-additive effects apparent in groups of the immune genes, we plotted expression levels for all immune genes in hybrids against parental mean expression of these genes, shown in Figure 3.3. Through these comparisons, it is evident that a substantial number of immune genes appear as outliers, indicative of non-additive expression in the hybrids. These outliers (arbitrarily defined as points that lie outside of the 95% confidence interval of the regression line) appear among genes expressed in uninfected flies, as well as in flies six or 12 hours after infection (Figure 3.3A). Strikingly, the patterns of outliers in expression levels vary widely among functional classes of immune genes. More specifically, genes coding for effectors are highly over-represented among the higher group of outliers (χ^2, d.f. = 1, $P = 7.19 \times 10^{-13}$), where hybrid expression is higher than parental mean expression, while genes coding for signaling proteins are highly over-represented among lower outliers (χ^2, d.f. = 1, $P = 4.03 \times 10^{-11}$), with lower hybrid than parental mean expression.

In addition to comparisons between hybrid and parental expression levels, we also examined levels of change in expression after infection in these flies (Figure 3.3B). Once again, it is apparent here that many genes in hybrids are induced (or repressed) at levels consistent with additivity, yet there are still numerous immune genes that appear as outliers in this comparison. In this case, there are not significant

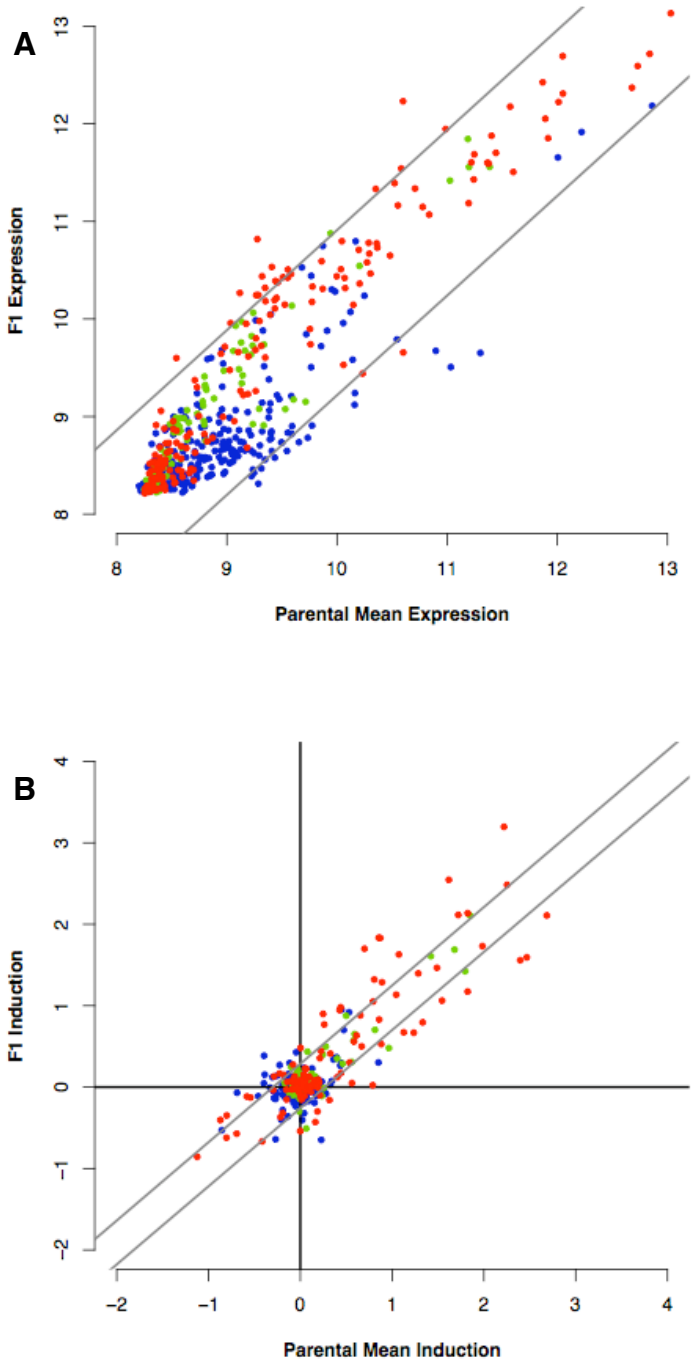


Figure 3.3 F1 vs. parental mean expression (A) and induction/repression (B) levels in immune genes. Recognition genes are shown in green, signaling genes in blue, and effectors in red. Gray lines represent 95% confidence intervals of the regression line for the data.

differences between outliers showing higher or lower induction levels, but effector genes are once again prominent; the genes from this functional class are significantly over-represented among all the outliers (χ^2, d.f. = 1, $P = 1.01 \times 10^{-9}$).

Coordinated Regulatory Differences in Hybrid Expression Change after Infection: To evaluate differences in induction or repression of immune genes in hybrids compared to flies from the parental species in the context of defined humoral pathways, we examined the patterns of genes clustered together based on similar patterns of induction/repression across samples (Figure 3.4). When the order of the clustered genes was compared to the order of gene products within the humoral immune pathways, we found these to be significantly correlated for genes in both the Toll and imd pathways (correlation coefficients = 0.643, 0.689; $P = 0.008, 0.017$, respectively). Through this, we see that induction patterns across all species groups appear to be more similar among genes closely grouped within the immune pathways.

To examine F1-specific patterns of coordinated gene regulation, we noted which uniquely induced or repressed genes in hybrids belonged to the Toll and imd pathways (Figure 3.5). Only a handful of genes throughout these pathways showed expression changes only in the hybrids, yet some patterns in the regulation of these genes are apparent, at least at six hours after infection. At this time point, the only genes that we observe to be uniquely regulated in the hybrids all belong to the imd pathway and show positive expression changes after infection, consistent with a systemic over-induction of genes in the imd pathway in response to Gram-negative infection in hybrids.

Bacterial Clearing Ability in Hybrids and Parental Species: As a proxy for systemic response to bacterial infection, we quantified bacterial load in terms of colony-forming units (cfu) per fly, at 12 and 25 hours after infection with *S. marcescens* in *D. melanogaster*, *D. simulans* and F1 hybrid flies. As shown in

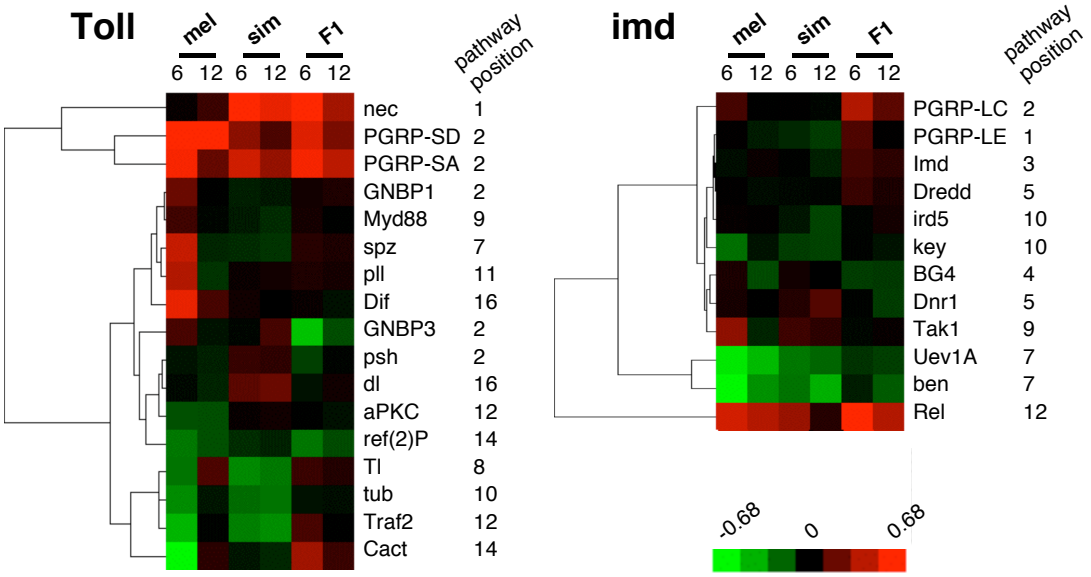


Figure 3.4 Genes throughout Toll and imd pathways (excluding effectors), clustered based on patterns of changes in expression levels after infection among samples. Scale bar indicates magnitude of expression change ($\log_2$(Signal) in infected flies – $\log_2$(Signal) in uninfected flies).

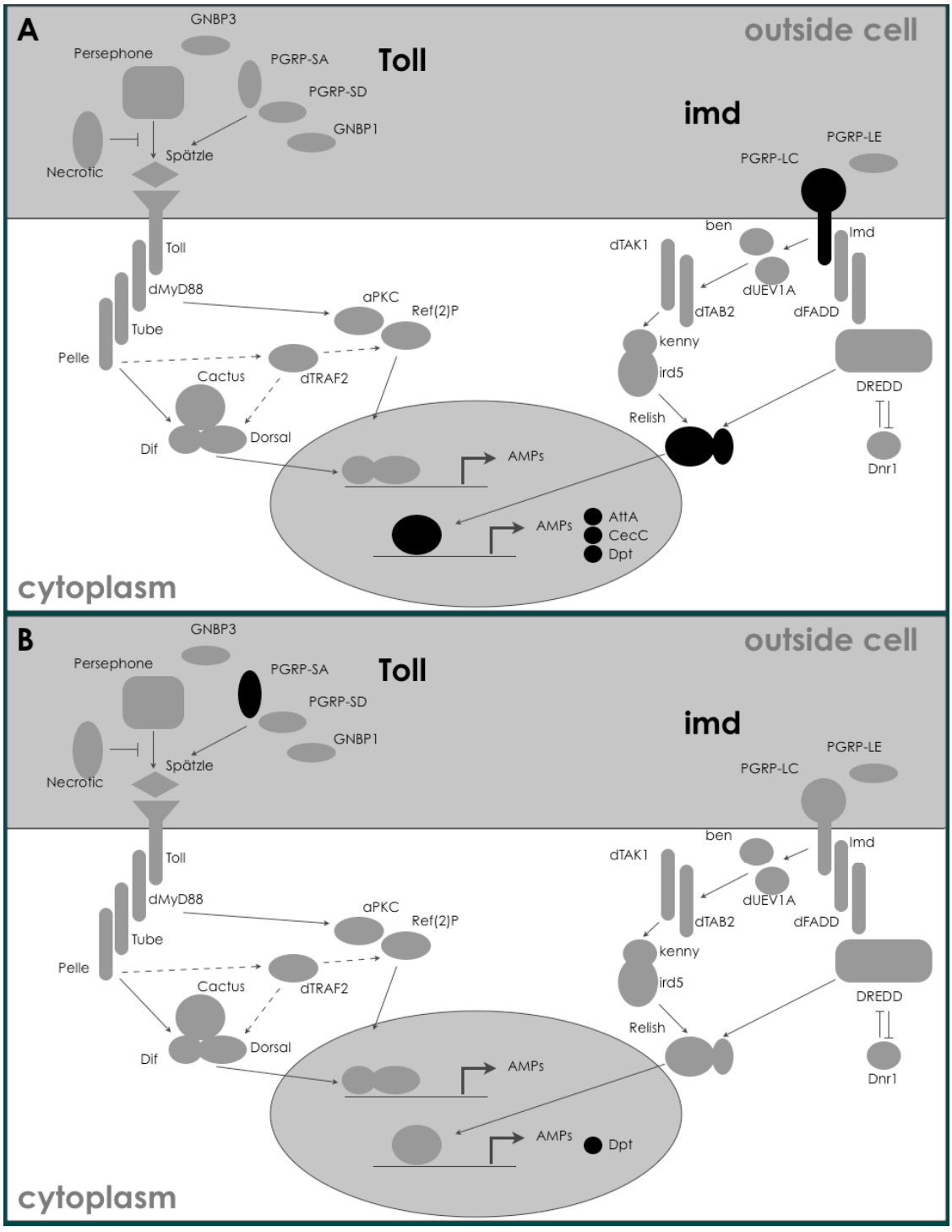


Figure 3.5 Genes uniquely induced in F1 hybrids throughout humoral immune pathways. Genes in black are induced only in hybrids six (A) or 12 hours (B) after infection with *S. marcescens*. Pathway genes and interactions included based on information in previous studies (ARBOUZOVA and ZEIDLER 2006; FERRANDON *et al.* 2007; FOLEY and O'FARRELL 2004; LECLERC and REICHHART 2004; STRONACH and PERRIMON 2002; WASSARMAN *et al.* 1995).

Figure 3.6, hybrid flies show similar bacterial levels after infection compared to the parental species, with no significant differences in load among groups of flies at 12 hours (ANOVA, $P = 0.9815$) or at 25 hours after infection (ANOVA, $P = 0.7719$). Additionally, mortality was observed in the flies with and without bacterial infection; the hybrids showed no significant difference in survival compared with the parental flies, up to 4 days after infection (*t-test*, $P = 0.251$).

Discussion

Here we report our findings of regulatory divergence in innate immunity between *D. melanogaster* and *D. simulans*, inferred from dysregulation of the immune response in the interspecific hybrids of these flies. Despite a general trend of additivity of expression of most immune genes in the hybrids, significant evidence for non-additive regulation was also detected across diverse parts of the innate immune pathway, indicative of divergent control of expression in response to bacterial infection in the two species. Interestingly, though the hybrids contain many genes that differ from parental levels of expression, the systemic immune phenotypes do not appear to be compromised in the hybrid flies; bacterial levels and survival rates after infection are consistent between hybrid and parental flies, indicating a robustness of the immune response to regulatory perturbations.

It is not surprising that the F1 hybrids examined here display distinctive expression profiles compared to parental species. If regulatory controls of transcription throughout the genome have diverged separately in two species, new combinations of *cis*- and *trans*-regulatory factors that arise in the hybrids may lead to unique expression patterns in the hybrids. Previous studies quantifying genome-wide transcript levels in closely related Drosophila species and their hybrids have also found substantial evidence for non-additive expression levels in interspecific hybrids

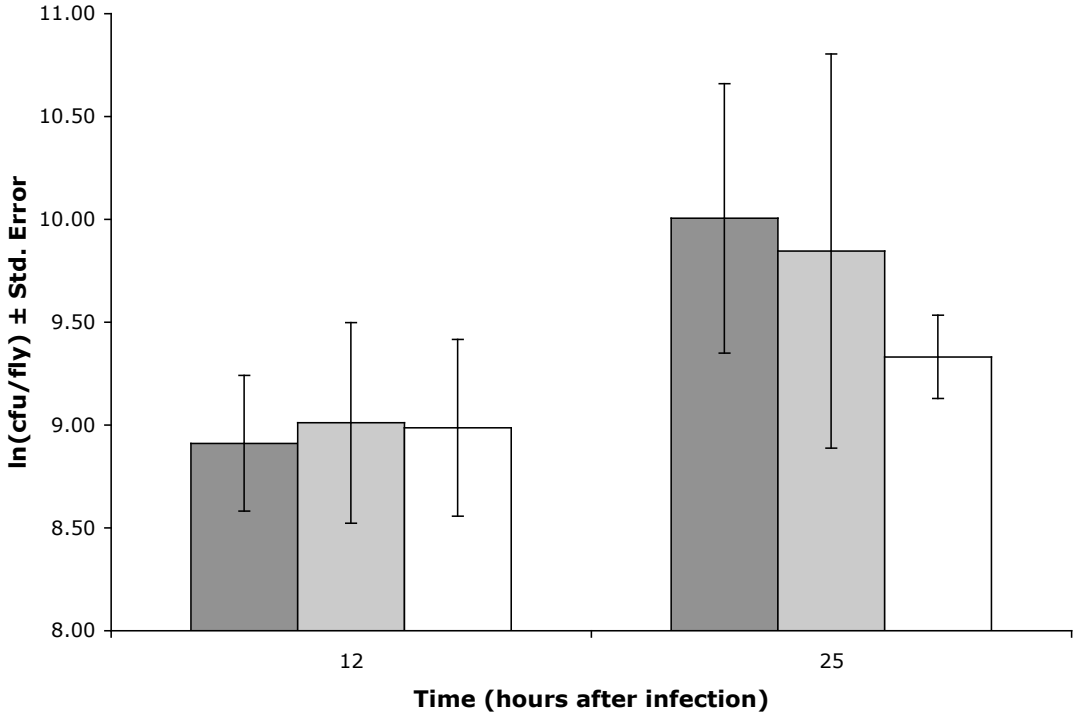


Figure 3.6 Bacterial load after infection in *D. melanogaster* (dark gray), *D. simulans* (white) and F1 hybrid (light gray) flies. Load represented as ln(cfu per fly) at 12 and 25 hours after infection with *S. marcescens*.

(RANZ *et al*. 2004). Using data from previous assays of genome-wide regulatory differences in interspecific Drosophila hybrids, ARTIERI *et al*. (2007) found significant correlations between sequence divergence between the parental species and hybrid dysregulation. We found no association between levels of amino acid divergence between *D. melanogaster* and *D. simulans* and levels of dysregulation in the hybrids as assessed by BeadChip analysis. We did, however, find a correlation between the difference in parental expression levels and degree of hybrid dysregulation for the genes examined (ANOVA, $P = 0.0218$). This may indicate that for the groups of genes examined here, the functional context and regulatory control of a gene may better predict hybrid dysregulation of the immune response than will interspecific divergence at the sequence level.

Beyond examining correlations of individual gene properties, such as sequence divergence, with that gene's expression in hybrid flies, we find that genes belonging to different functional groups within the innate immune response show distinct patterns of dysregulation in hybrids. These functional groups, bearing genes whose products are involved in recognition, signaling, and effector roles in the response to infection, have been previously shown to have distinct patterns of sequence diversity – in terms of both naturally occurring variation within populations as well as divergence among species (LAZZARO *et al*. 2006; LAZZARO *et al*. 2004). The hybrid dysregulation that we see here indicates that interspecific divergence in immune pathway genes manifests differently not only in sequence differences, but also in distinct regulatory patterns among groups of immune genes. In this case, the effector genes appear to be the most dysregulated in hybrids after infection; these tend to be over-expressed in the hybrids at all timepoints, and they are the group of genes most likely to show aberrant levels of induction or repression in the hybrids after infection. The patterns of disrupted expression in these downstream components of the immune response could indicate

that regulatory differences in upstream components propagate throughout the network, leading to the highest levels of dysregulation in these effectors.

In addition to being the group with an abundance of genes displaying unique patterns of expression in the hybrids, the effectors also show greater levels of hybrid overexpression than other functional groups. This directionality of the disruption may indicate that the regulatory breakdown in these hybrids is not random. We might expect some level of overexpression of immune genes in the hybrid flies relative to parents, since hybrids between *D. melanogaster* and *D. simulans* have been shown to have enlarged fat bodies (DICKINSON *et al*. 1984); however, this would not necessarily lead to distinct differences in hybrid dysregulation among the genes corresponding to separate functional groups within the immune pathways. The widespread overexpression of the downstream genes in the hybrid immune response may reflect some sort of basal regulatory mechanism of the pathways; perhaps the immune system in these flies is primed to have high AMP expression in response to infection.

While an active response to infection may be beneficial in the face of regular exposure to microbes, it could be costly if it takes too much energy away from other vital tasks. To counter this, the Drosophila immune response typically includes negative feedback mechanisms. Since the hybrid flies in this study display a more active response to infection than the parental flies, this may represent a disruption of the negative feedback in the immune pathways, potentially due to divergent regulation of these mechanisms between the two parental species. If such disruption is present, we would expect there to be decreased expression levels of genes with roles in repression of the immune response in hybrids. We do, in fact, find evidence for this – we observe that *Dnr1*, previously shown to repress activation of the imd pathway (FOLEY and O'FARRELL 2004), is downregulated in hybrids more than in parents at both six and 12 hours after infection, though most of these differences are not

significant. It should also be noted that not all genes classified as members of the group of effectors are antimicrobial peptides. Some of these genes encode proteins involved in stress response that can be transcribed in response to infection, yet may be induced as a result of other stresses – either at a systemic or cellular level. While overexpression of these genes may reflect an overactive immune system, they could also reflect instances of dysregulation in the hybrids independent of the response to infection.

While differences in the regulation of immune gene expression are evident between the parental species and with the F1 hybrids, it is also apparent that induction and repression patterns are at least somewhat consistent among genes that are closely positioned functionally in the humoral immune pathways. This is not entirely expected, since the Toll and imd pathways are not transcriptional networks; upstream genes in the pathways do not directly control the transcription of their downstream neighbors. On the other hand, it is not surprising that there would be some level of coordination of expression of genes clustered within the pathways, since the gene products of these genes interact. This does show that not only the expression changes after infection, but also their relative magnitudes across genotypes, are somewhat consistent with their positions in the humoral pathways, implying coordinated regulation of expression as well as evolution of that regulation.

The observation that members of the Toll pathway are dysregulated in hybrids also raises questions about the effects of imprecise regulation of pleiotropic genes. With its role in the control of embryonic development as well as immune response in Drosophila, an intact Toll pathway is critical for proper function of multiple phenotypes at various life stages in the fly. Patterns of dysregulation of Toll genes could be associated with developmental differences in hybrids; distinct morphological and developmental features have in fact been observed in hybrids of these Drosophila

species (DAVID *et al*. 2002; MARKOW and RICKER 1991; STURTEVANT 1920). Dysregulation of the Toll pathway in these interspecific hybrids has clearly extended beyond control of embryonic development, though; differences in induction and repression of Toll genes in hybrids as a result of bacterial infection in adult flies are indicative of regulatory divergence specific to the immune response.

As pairs of species diverge, the hybrids that they may form can bear dysfunctional phenotypes, ranging from inviability and sterility to more subtle differences in morphology or regulation. Through the juxtaposition of two diverged genomes, it is likely that numerous systems throughout a hybrid individual may be disrupted and that even seemingly subtle regulatory differences could have fitness consequences (ORTIZ-BARRIENTOS *et al*. 2007). In this investigation of immune dysregulation of interspecific Drosophila hybrids, we have found that regulatory divergence of the innate immune system between *D. melanogaster* and *D. simulans* manifests distinctly in different portions of the immune response in F1 hybrids, with the most notable disparities appearing in the downstream pathway components. Furthermore, despite clear differences in patterns of expression and induction after infection in the F1 hybrids, these flies appear as immunocompetent as flies from the parental species, revealing a robustness of the immune function to even widespread regulatory perturbations, and potentially highlighting an evolved ability of the immune networks to tolerate expression differences – at least in flies maintained in laboratory conditions.

REFERENCES

APIDIANAKIS, Y., M. N. MINDRINOS, W. XIAO, G. W. LAU, R. L. BALDINI *et al.*, 2005 Profiling early infection responses: *Pseudomonas aeruginosa* eludes host defenses by suppressing antimicrobial peptide gene expression. Proc Natl Acad Sci U S A **102:** 2573-2578.

ARTIERI, C. G., W. HAERTY and R. S. SINGH, 2007 Association between levels of coding sequence divergence and gene misregulation in Drosophila male hybrids. J Mol Evol **65:** 697-704.

AUGER, D. L., A. D. GRAY, T. S. REAM, A. KATO, E. H. COE, JR. *et al.*, 2005 Nonadditive gene expression in diploid and triploid hybrids of maize. Genetics **169:** 389-397.

CLARK, A. G., and L. WANG, 1997 Molecular population genetics of Drosophila immune system genes. Genetics **147:** 713-724.

DATE, A., Y. SATTA, N. TAKAHATA and S. I. CHIGUSA, 1998 Evolutionary history and mechanism of the Drosophila *cecropin* gene family. Immunogenetics **47:** 417-429.

DAVID, J. R., P. GIBERT, G. PETAVY and B. MORETEAU, 2002 Variable modes of inheritance of morphometrical traits in hybrids between *Drosophila melanogaster* and *Drosophila simulans*. Proc Biol Sci **269:** 127-135.

DE GREGORIO, E., P. T. SPELLMAN, G. M. RUBEN and B. LEMAITRE, 2001 Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. Proc Natl Acad Sci U S A **98:** 12590-12595.

DE HOON, M. J. L., S. IMOTO, J. NOLAN and S. MIYANO, 2004 Open Source Clustering Software. Bioinformatics **20:** 1453-1454.

DICKINSON, W. J., R. G. ROWAN and M. D. BRENNAN, 1984 Regulatory Gene Evolution: Adaptive Differences of Alcohol Dehydrogenase in *Drosophila melanogaster* and *Drosophila simulans*. J. Hered. **52:** 215-225.

DOVER, G. A., 1992 Observing development through evolutionary eyes: a practical approach. Bioessays **14:** 281-287.

FERRANDON, D., J. L. IMLER, C. HETRU and J. A. HOFFMANN, 2007 The Drosophila systemic immune response: sensing and signalling during bacterial and fungal infections. Nat Rev Immunol **7:** 862-874.

FOLEY, E., and P. H. O'FARRELL, 2004 Functional dissection of an innate immune response by a genome-wide RNAi screen. PLoS Biol **2:** E203.

IRVING, P., L. TROXLER, T. S. HEUER, M. BELVIN, C. KOPCZYNSKI *et al.*, 2001 A genome-wide analysis of immune responses in Drosophila. Proc Natl Acad Sci U S A **98:** 15119-15124.

LANDRY, C. R., D. L. HARTL and J. M. RANZ, 2007 Genome clashes in hybrids: insights from gene expression. Heredity **99:** 483-493.

LAZZARO, B. P., 2005 Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans*. Genetics **169:** 2023-2034.

LAZZARO, B. P., and A. G. CLARK, 2003 Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. Mol Biol Evol **20:** 914-923.

LAZZARO, B. P., T. B. SACKTON and A. G. CLARK, 2006 Genetic variation in *Drosophila melanogaster* resistance to infection: a comparison across bacteria. Genetics **174:** 1539-1554.

LAZZARO, B. P., B. K. SCEURMAN and A. G. CLARK, 2004 Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. Science **303:** 1873-1876.

LECLERC, V., and J. M. REICHHART, 2004 The immune response of *Drosophila melanogaster*. Immunol Rev **198:** 59-71.

MARKOW, T. A., and J. P. RICKER, 1991 Developmental stability in hybrids between the sibling species pair, *Drosophila melanogaster* and *Drosophila simulans*. Genetica **84:** 115-121.

MEDZHITOV, R., and C. A. JANEWAY, JR., 1997 Innate immunity: impact on the adaptive immune response. Curr Opin Immunol **9:** 4-9.

MICHALAK, P., and M. A. NOOR, 2003 Genome-wide patterns of expression in Drosophila pure species and hybrid males. Mol Biol Evol **20:** 1070-1076.

MOEHRING, A. J., K. C. TEETER and M. A. NOOR, 2007 Genome-wide patterns of expression in Drosophila pure species and hybrid males. II. Examination of multiple-species hybridizations, platforms, and life cycle stages. Mol Biol Evol **24:** 137-145.

ORTIZ-BARRIENTOS, D., B. A. COUNTERMAN and M. A. NOOR, 2007 Gene expression divergence and the origin of hybrid dysfunctions. Genetica **129:** 71-81.

PARKER, H. R., D. P. PHILIPP and G. S. WHITT, 1985 Gene regulatory divergence among species estimated by altered developmental patterns in interspecific hybrids. Mol Biol Evol **2:** 217-250.

RAMOS-ONSINS, S., and M. AGUADE, 1998 Molecular evolution of the Cecropin multigene family in Drosophila. functional genes vs. pseudogenes. Genetics **150:** 157-171.

RANZ, J. M., K. NAMGYAL, G. GIBSON and D. L. HARTL, 2004 Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. Genome Res **14:** 373-379.

SACKTON, T. B., B. P. LAZZARO, T. A. SCHLENKE, J. D. EVANS, D. HULTMARK *et al.*, 2007 Dynamic evolution of the innate immune system in Drosophila. Nat Genet **39:** 1461-1468.

SALDANHA, A. J., 2004 Java Treeview--extensible visualization of microarray data. Bioinformatics **20:** 3246-3248.

SCHLENKE, T. A., and D. J. BEGUN, 2003 Natural selection drives Drosophila immune system evolution. Genetics **164:** 1471-1480.

SCHMID-HEMPEL, P., 2003 Variation in immune defence as a question of evolutionary ecology. Proc Biol Sci **270:** 357-366.

SHAW, P. J., N. S. WRATTEN, A. P. MCGREGOR and G. A. DOVER, 2002 Coevolution in bicoid-dependent promoters and the inception of regulatory incompatibilities among species of higher Diptera. Evol Dev **4:** 265-277.

STURTEVANT, A. H., 1920 Genetic Studies on *DROSOPHILA SIMULANS*. I. Introduction. Hybrids with *DROSOPHILA MELANOGASTER*. Genetics **5:** 488-500.

TRUE, J. R., and E. S. HAAG, 2001 Developmental system drift and flexibility in evolutionary trajectories. Evol Dev **3:** 109-119.

WHITT, G. S., W. F. CHILDERS and P. L. CHO, 1973 Allelic Expression at Enzyme Loci in an Intertribal Hybrid Sunfish. J. Hered. **64:** 55-61.

WHITT, G. S., D. P. PHILIPP and W. F. CHILDERS, 1977 Aberrant Gene Expression during the Development of Hybrid Sunfishes (Perciformes, Teleostei). Differentiation **9:** 97-109.

CHAPTER 4

GENOME-WIDE REGULATORY DIVERGENCE AFTER BACTERIAL INFECTION IN INTERSPECIFIC DROSOPHILA HYBRIDS

Abstract

In order to assess the divergence of the innate immune response and genome-wide consequences of bacterial infection between Drosophila species, we have quantified transcript abundance before and after infection in *D. melanogaster*, *D. simulans* and their F1 hybrids using Illumina short-read sequencing. If pathways in this system have diverged such that incompatibilities have arisen between interacting components of the immune network, we expect the hybrids to display dysregulation of immune genes. Furthermore, since immune system activation may impact the entire organism, we have examined the relationship between regulatory divergence and the influences of a systemic response to infection by assaying genome-wide transcript levels in these flies. With this, we find evidence for regulatory differences in the hybrids relative to the parental species after infection. Consistent with our previous study (Chapter 3), we observe distinct regulatory patterns among separate functional groups of genes within the immune response, and we also see hybrid dysregulation in non-immune pathways in response to infection, indicative of the broad effects of divergence in the innate immune response.

Introduction

As closely related species evolve, they tend to display increasingly divergent genotypes and phenotypes. If these species are relatively recently diverged, however, some differences may be subtle, and changes in some genes may be balanced by compensatory changes in other genes, yielding phenotypes that are not obviously

different between the species. Some divergence at this level can be uncovered by examining F1 hybrids between the diverged pair of species. Since interspecific hybrids bear two distinct genomes juxtaposed in a single individual, they have the potential to display dysregulation of the parental phenotypes if diverged regulatory mechanisms do not act appropriately when combined (LANDRY *et al.* 2007; TRUE and HAAG 2001). Previous studies have uncovered evidence of divergence of key functions within hybrid individuals, leading to differences in enzyme levels or developmental patterns in these organisms, when compared to the parental species (DICKINSON *et al.* 1984; PARKER *et al.* 1985; WHITT *et al.* 1977).

More recently, studies of hybrid dysregulation have expanded to the genomic scale as a variety of resources have become available for studying genome-wide expression levels in numerous species. While some of these studies have assayed regulatory divergence in taxa such as whitefish, maize, and other plants (AUGER *et al.* 2005; HEGARTY *et al.* 2005; RENAULT *et al.* 2009; SWANSON-WAGNER *et al.* 2006), a concerted effort has been focused on expression patterns in interspecific Drosophila hybrids (MICHALAK and NOOR 2003; MOEHRING *et al.* 2007; RANZ *et al.* 2004; REILAND and NOOR 2002). One common finding among many of these studies is that hybrids tend to display non-additive expression for many genes throughout the genome. This may indicate that as species diverge, regulation of expression evolves throughout the genome, and that these regulatory controls interact such that the hybrid expression is driven outside of the parental range.

Beyond constitutive transcript abundance and developmental or reproductive defects, it is expected that dysregulation in hybrids would be likely to occur in regulatory networks throughout the genome, and that responses to outside stimuli by such networks may be compromised in hybrids, or at least improperly activated. One such network is that of the innate immune response in Drosophila. A complex set of

pathways, involving recognition of invading microbes, signaling transduction, and the transcription of effectors such as antimicrobial peptides (AMPs), the innate immune network would seem susceptible to regulatory divergence between species leading to improper immune response in interspecific hybrids. Previously, we found evidence for such dysregulation when we quantified expression levels before and after infection in a set of immune and non-immune genes (Chapter 3). Through this, we found non-additive expression patterns, as well as overactivation of portions of the immune response in hybrids, with variation in this dysregulation among genes belonging to separate functional groups within the immune response. This is consistent with previous findings that genes from these groups show distinct levels of polymorphism and divergence among Drosophila species (SACKTON *et al.* 2007).

Here, we have further investigated the level of regulatory divergence in immune response in Drosophila by assaying genome-wide transcript abundance in F1 hybrids using Illumina short-read sequencing of cDNA before and after infection. This sequencing method (RNA-seq) has been gaining popularity as a flexible approach to quantify genome-wide expression levels, and has been used successfully in a number of model systems so far (CLOONAN *et al.* 2008; MORTAZAVI *et al.* 2008; NAGALAKSHMI *et al.* 2008). Not only does this allow us to query transcript levels for genes across the genome to evaluate widespread impacts of bacterial infection in the hybrids, but we can also distinguish parental alleles within the F1 hybrids to quantify allele-specific dysregulation in these flies. With this, we have found evidence for hybrid dysregulation in the response to infection that varies among different functional components of the immune response, consistent with our previous findings of distinct patterns of evolution among these groups. Furthermore, we have found dysregulation of other pathways throughout the genome, potentially indicative of a cost of immune activation that may also be affected by divergent regulatory patterns in the hybrids.

Materials and Methods

Fly Lines and Crosses: We used inbred laboratory stocks of the *D. melanogaster zygotic hybrid rescue* (*zhr*) strain (provided by A. Orr) and a Tsimbazaza strain of *D. simulans* (provided by H. Hollocher) for the hybrid crosses. We collected virgin *D. melanogaster* females and *D. simulans* males and mated the flies in vials with approximately 10 females and 10 males apiece. Crosses including only *D. melanogaster* or only *D. simulans* males and females were also set up, to yield flies from the parental species reared under similar conditions. We collected F1 hybrid female flies (since male hybrids from this cross are not viable), as well as females from each of the intraspecific crosses, for infection.

Bacterial Cultures and Infections: To provoke a response involving multiple aspects of the innate immune pathways, we infected flies with both Gram-positive and Gram-negative bacteria simultaneously. To achieve this, we used a combination of *Enterococcus faecalis* & *Serratia marcescens*. Cultures of Gram-positive *E. faecalis* (derived from the strain used by LAZZARO *et al.* (2006)), identified using 16S rDNA sequence and the results of API 20Strep substrate utilization) and Gram-negative *S. marcescens*, derived from ATCC strain 13880 (also previously used by LAZZARO *et al.* (2004)) were combined at a 1:1 ratio after being separately grown overnight from freezer stocks and diluted to a concentration with $OD_{600} \sim 1.0$. We infected flies by dipping 0.1-mm tungsten needles (Fine Science Tools, Foster City, CA) into the bacterial mix and using these to pierce the flies' thoraces. We froze both infected and uninfected flies 12 hours later, using liquid nitrogen.

Transcript Quantification via Short-Read Sequencing: We isolated mRNA from whole flies (approximately 50 females per sample) using a Trizol:chloroform extraction and then synthesized cDNA using oligo-dT priming. cDNA transcripts

were sequenced using the Illumina Solexa GA2 platform at the Cornell Bioresource Center. One lane was used to sequence each of the six samples (*D. melanogaster*, *D. simulans* and F1 hybrid flies in either uninfected or infected states), providing approximately 300-400 megabases of sequence for each, with 36 bases per sequenced read.

Alignment of Reads to Reference Genomes: We processed sequenced reads and aligned them to reference sequences using Maq (Mapping and Assembly with Qualities) software (LI *et al.* 2008). To identify reads corresponding to transcripts throughout the genomes of both species, we used sequences representing the *D. melanogaster* or *D. simulans* genome, from genome-wide alignments for *D. melanogaster* and *D. simulans* (alignments produced using MAVID (BRAY and PACHTER 2004), obtained from Assembly/Alignment/Annotation of 12 related Drosophila species website, http://rana.lbl.gov/drosophila/). These references included 76.7% and 93.9% of the *D. melanogaster* and *D. simulans* complete genome sequence available, respectively.

Using Maq protocols, we converted Illumina output sequences and reference sequences to the appropriate formats for alignment. Reads from each sample were aligned to the appropriate reference genome(s), with the Maq settings tolerating the alignment of reads bearing up to three mismatches with the reference. Sequences from *D. melanogaster* and *D. simulans* samples were aligned to their respective genomes, and the sequences from F1 hybrid samples were aligned to both references. This yielded matches for 23.4 to 34.8% of the reads for each sample. Only hits with quality scores of 10 or greater were analyzed, eliminating any non-unique alignments genome-wide.

Quantification and Normalization of Read Density: To examine transcript levels for each gene in the genome, we quantified read density for each sample at

bases annotated in the *D. melanogaster* genome. For each base annotated as uniquely belonging to the transcript of one gene, we classified the base as either being identical between *D. melanogaster* and *D. simulans*, divergent between the species, or as a gap in the alignment. At each identical base within a transcript, we counted the number of hits that aligned for a given sample. For total transcript abundance in the hybrids, we only used the alignment to the *D. melanogaster* reference genome. Since only non-divergent sites were used to determine transcript levels, both parental alleles should be detected here. Furthermore, we expect that a high proportion of alleles from the *D. simulans* parent will still align to a *D. melanogaster* reference sequence; when reads sequenced from the *D. simulans* sample were aligned to the *D. melanogaster* reference as a test, 85% as many reads matched as when the same reads were aligned to the *D. simulans* genome.

To normalize the read counts across samples, we divided counts in each sample by the total number of mapped reads in that sample (with each sample multiplied by the maximum number of aligned reads among samples to keep the numbers at the same magnitude). To control for different reference lengths among genes, we divided the normalized count totals for each transcript by the number of identical sites in that transcript, yielding a read density (number of hits per kb) for each gene in the genome. This normalization is similar to that used by MORTAZAVI *et al.* (2008) for quantifying transcript levels through RNA-seq.

Allele-specific Read Density in Hybrids: Since interspecific hybrids bear alleles from two parental genomes, we also wanted to assay transcript abundance from each parental allele in the F1 hybrid flies. For this, we aligned sequences from F1 hybrid samples to both *D. melanogaster* and *D. simulans* reference genomes and counted the number of hits belonging to each parental allele at each divergent site within annotated transcripts. Out of all divergent sites within annotated transcripts,

only those with at least 5 hits in one or more samples were included. These counts were also normalized across samples, and the total number of counts at all divergent sites within a gene was normalized by the length of the transcript.

Statistical Analysis: To test for effects of infection and genotype × infection interactions on phenotypes, we performed ANOVAs on the data, for groups of genes categorized by functional pathways. Pathways were determined based on annotations from the Kyoto Encyclopedia of Genes and Genomes (KEGG, KANEHISA and GOTO 2000). For this, we tested the following model for groups of transcribed genes belonging to each pathway:

$$y_{ijk} = \mu + \text{Infection}_i + \text{Genotype}_j + (\text{Infection} \times \text{Genotype})_k + \varepsilon_{ijkl} \qquad (1)$$

where y represents the phenotype of interest for each gene in the group (expression levels or proportion of parental alleles), and Infection (i = 1,2), Genotype (j = 1,2), and their interaction term are fixed effects. Since pathway members were not entirely independent (some transcribed genes belonged to more than one annotated pathway), we determined the significance of each effect by permuting the data 1,000 times and comparing the actual statistic for each to the null distribution of statistics from the permuted data.

Results

Transcripts Identified by Short-Read Sequencing: After aligning sequenced reads from each sample to the reference genomes, transcripts from many of the annotated genes in the *Drosophila melanogaster* genome were detected. Out of 13,400 annotated genes that included sites aligned between *D. melanogaster* and *D. simulans*, 10,977 (81.9%) of these had sites to which sequenced reads aligned uniquely. Of

these, we have defined 4,574 as being transcribed in one or more samples, using a read density cutoff of at least 36 hits, or one fully sequenced read, per kb. Many of these appear to be commonly transcribed in *D. melanogaster*, and *D. simulans*, but not necessarily in F1 hybrid flies, with 1,253 (27.4%) of the transcribed genes appearing in the two parental species, and only 693 (15.2%) transcribed in all three groups. Figure 4.1 illustrates the overlap of the transcribed genes among these genotypes.

Transcript Abundance in Parental and Hybrid Flies: Transcript abundance in hybrids is positively correlated with the parental mean transcript levels genome-wide (Pearson's correlation, $P < 2.2 \times 10^{-16}$), indicative of largely additive effects acting upon expression levels in the hybrid flies. Figure 4.2A shows a plot of these transcript levels, represented as $\log_2$(hits per kb). We also see, though, that expression levels in hybrids are much more variable than the parental mean levels, with significant differences in variance between the groups both before and after infection (F-test, $P < 1.0 \times 10^{-30}$, $P = 3.22 \times 10^{-49}$, respectively). Furthermore, differences in expression between hybrids and parents vary widely among genes; some genes show up to 655-fold higher transcript abundance in the parents, while others have up to 864-fold higher levels of transcript in the hybrid flies. Figure 4.2B illustrates the distribution of these differences among transcribed genes, before and after infection.

To evaluate the effects of bacterial infection on genome-wide expression levels in hybrids, we also examined the changes in transcript levels before and after infection in hybrids and parental species. As with the actual transcript levels, the levels of induction or repression show varying levels of similarity between hybrids and parents. Figure 4.3 shows changes in transcript levels in hybrids relative to those in parents,

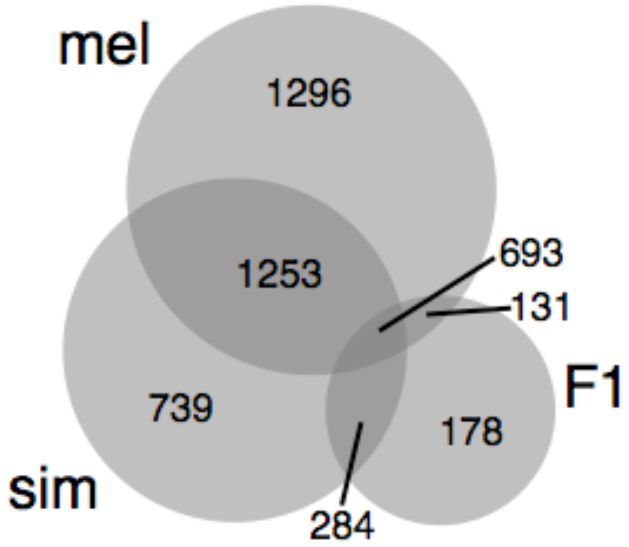


Figure 4.1 Counts of genes transcribed in one or more of the groups of flies (with at least 36 hits per kb), in either uninfected or infected samples.

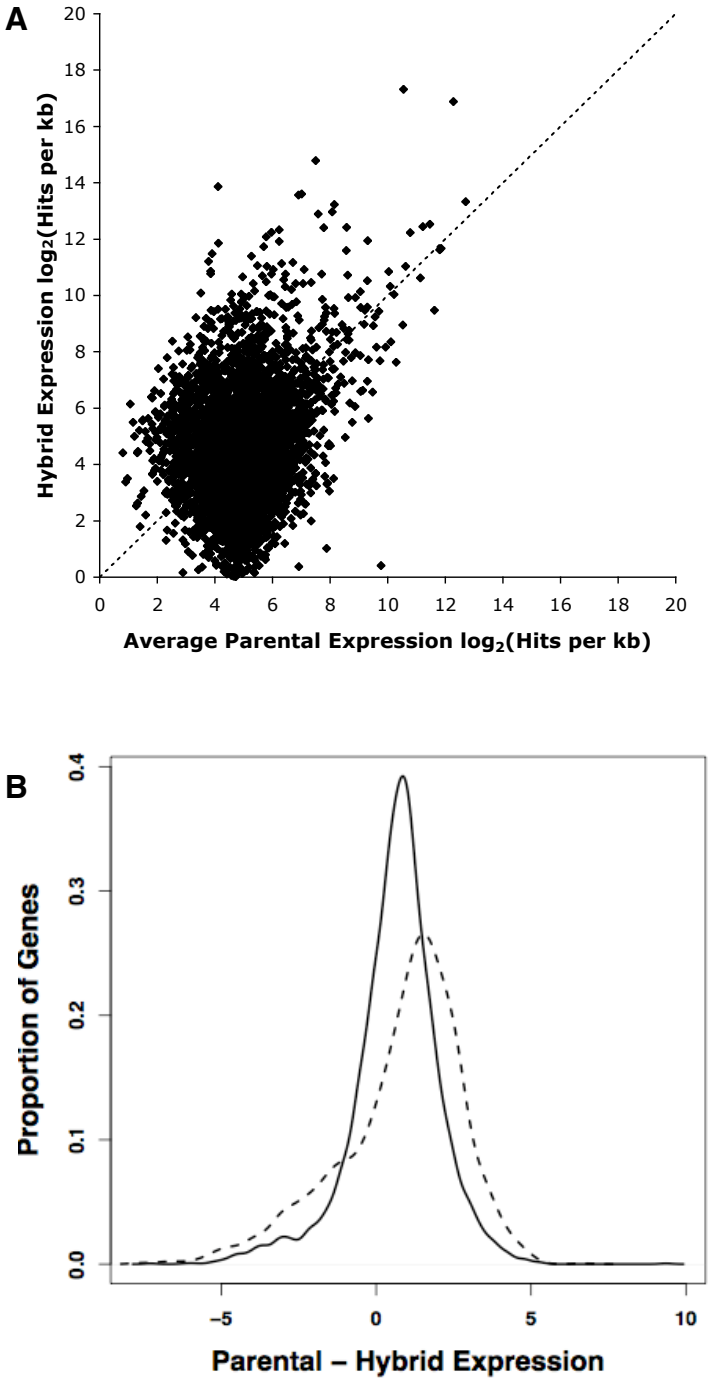


Figure 4.2 (A) Genome-wide transcript levels in hybrids and parents, as $\log_2$(Hits/kb), for genes before and after bacterial infection. Dotted line corresponds to y = x. (B) Distribution of differences in parental and hybrid expression levels (parental mean – hybrid $\log_2$(Hits/kb)) for transcribed genes throughout the genome, in uninfected flies (dashed line) and infected flies (solid line).

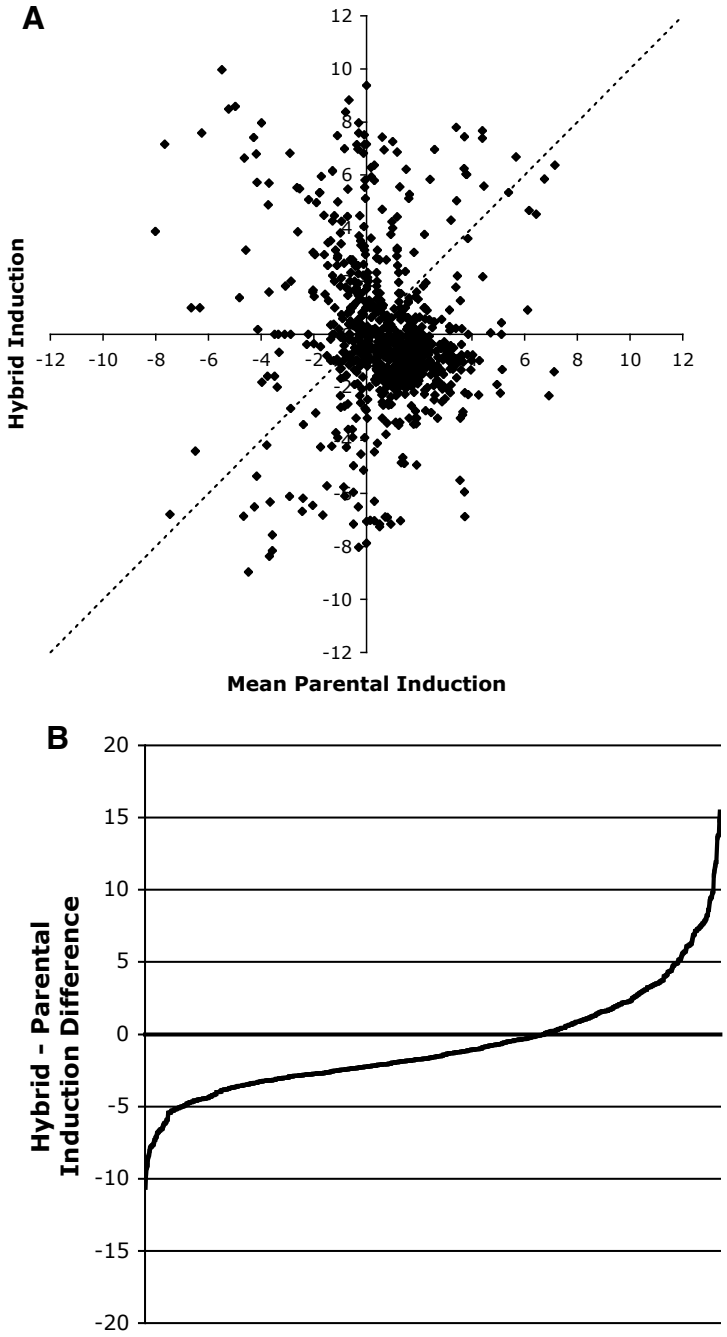


Figure 4.3 Changes in transcript level after infection in hybrids and parental species (A) and disparities between induction levels in the two groups (B) among genes with at least 100 hits per kb in one or more samples. Dashed line in (A) corresponds to y = x.

among the most highly transcribed genes (those with at least 100 hits per kb in one or more samples). Interestingly, we observe more genes with greater induction in parents, yet genes showing the greatest difference between genotypes are induced more highly in hybrids.

Functional Differences in Hybrid Dysregulation: While it is apparent that F1 hybrids can show distinct regulation of expression in response to infection compared with the parental species, it has also become clear that these patterns of hybrid dysregulation are not consistent among groups of genes with different functional roles. To directly assay the response to infection in these interspecific hybrids, we examined the transcription of immune-related genes before and after infection. Overall, genes belonging to the innate immune pathways show a significant increase in expression level after infection (ANOVA, $P < 2.20 \times 10^{-16}$); different functional groups within the immune response show distinct patterns, though. Figure 4.4A illustrates mean expression levels for groups of genes coding for recognition, signaling and effector proteins in hybrids and parents before and after infection (transcribed genes included in each functional group listed in Table C.1). While genes in each of these groups show significant increases in expression after infection, genes in the effector class show a distinct pattern of expression, with generally higher transcript levels, relative to other immune genes. Effectors also show fairly high levels of variance in expression level in each sample, likely due to a subset of genes being highly expressed following infection, while others are not activated – genes in the effector class are not all antimicrobial peptides, so it is unlikely that all genes in this functional group would be identically regulated after infection. Additionally, genes in the signaling group show a significant effect of the interaction between genotype and infection; these genes are more highly induced in hybrids than in the parents.

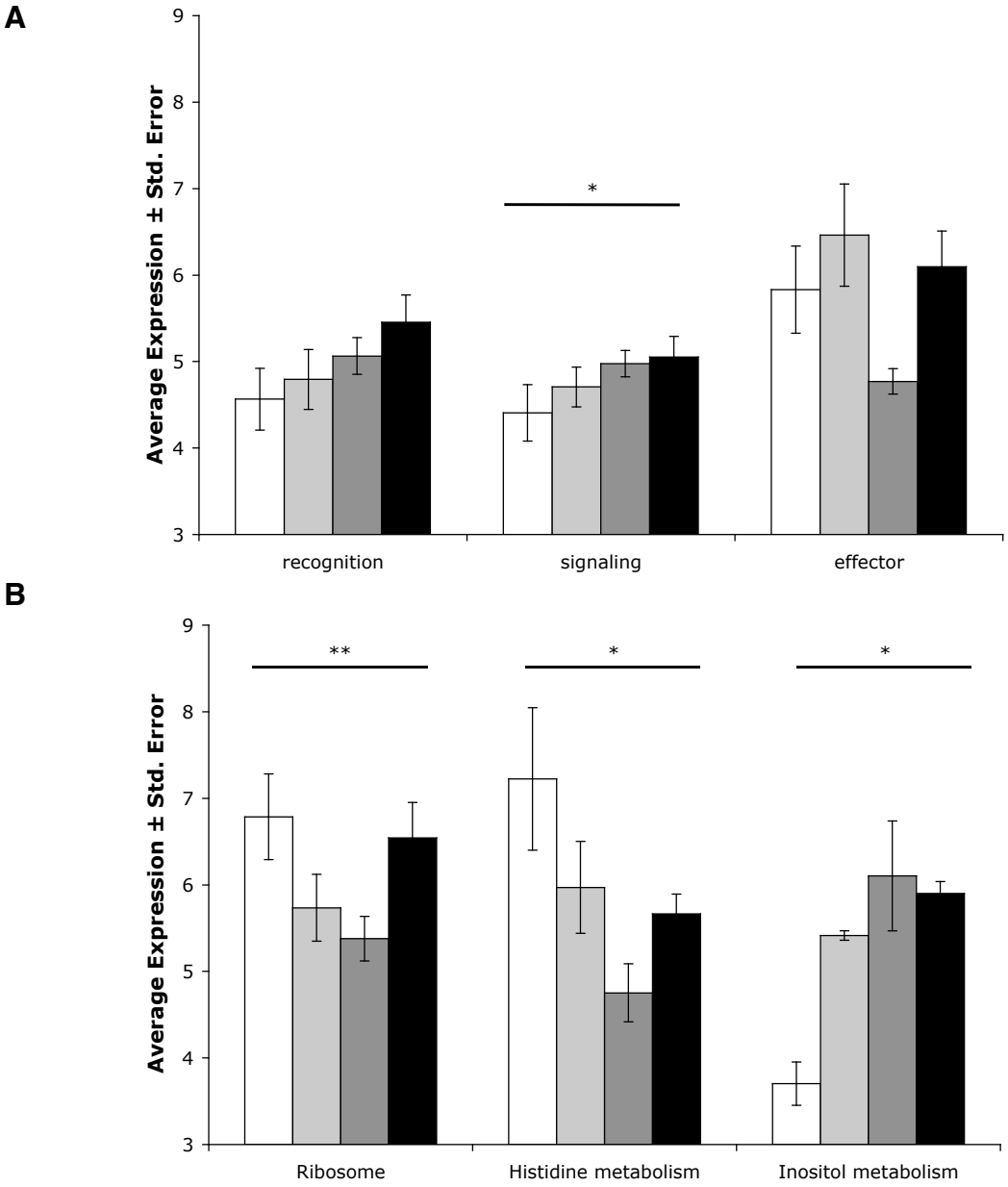


Figure 4.4 Expression levels before and after bacterial infection among genes in different functional groups within the immune pathways (A) or involved in other pathways throughout the genome (B). Hybrid uninfected and infected expression levels (white and light gray bars) and parental mean uninfected and infected expression levels (dark gray and black bars) are shown, as the average of $\log_2$(hits per kb) for each group. Non-immune pathways are included for which there is a significant effect of a genotype x infection interaction ($P < 0.05$) on expression. Asterisks in each plot indicate genotype x infection interactions with * $P < 0.05$ and ** $P < 0.01$.

Beyond the defined members of the innate immune response, we expect other genes throughout the genome to undergo expression changes after infection, either as a direct reaction to bacterial infection or as a consequence of the impact of immune response on the rest of the organism. Furthermore, if non-immune genes play a critical role after infection, and if this function is disrupted in interspecific hybrids, we expect hybrids to induce or repress these genes at a different magnitude than the parental species. To investigate coordinated non-immune responses to infection, we tested expression difference in hybrids and parents before and after infection in other pathways throughout the genome, as identified through KEGG. Table 4.1 lists pathways with a significant effect of either infection or a genotype by infection interaction on expression levels (based on ANOVA tests, with $P < 0.05$). For the three pathways here affected by a genotype-by-infection interaction, we have plotted expression means in Figure 4.4B. Notably, the non-immune pathways with the most significant interactions show decreased expression in hybrids, while the parents show induction in the included genes after infection. Furthermore, most of the non-immune pathways with apparent changes in expression after bacterial infection appear related to metabolism and other housekeeping processes. This could be indicative of indirect effects or costs on the system as a result of immune response, some of which may be sensitive to the genotype of two diverged parental alleles in these flies.

Allele-specific Differences in Transcript Abundance in Hybrids: In addition to overall transcript levels in hybrids relative to flies of the parental species, we also examined the relative abundances of the two parental alleles in the hybrids before and after infection. If expression regulation of a given gene has evolved between *D. melanogaster* and *D. simulans*, we expect that at least some of that expression difference may be reflected in allele-specific transcript abundance in the F1 hybrids. Among related groups of transcribed genes, if the *D. melanogaster* allele

Table 4.1 Pathways Showing Significant Effects of Infection on Expression

Pathway Name	Infection	Genotype x Infection
Ribosome	0.075	**0.003**
Histidine metabolism	0.107	**0.037**
Inositol metabolism	**0.028**	**0.048**
Bisphenol A degradation	**0.038**	0.051
Bile acid biosynthesis	**0.037**	0.059
Biosynthesis of unsaturated fatty acids	**0.048**	0.106
DNA replication	**0.042**	0.656

ANOVA *P* -values < 0.05 indicated in bold

comprises approximately 50% of the transcript for all genes (yielding an average proportion for the group of about 0.5, with a low level of variance), we assume that the transcripts for genes in that group in hybrids are fairly equally represented, and that expression is mostly additive. Conversely, if genes in a group are consistently regulated with a bias toward the expression of one parental allele over the other, we expect to observe the average proportion of the *D. melanogaster* allele for the group to be either above or below 0.5, but the variance should still be low. Furthermore, if the allelic proportions within a group have a wide variance, the genes in that group appear to be less coordinated with regard to allele-specific transcription.

To look at allelic transcription levels within the immune response and their changes after infection, we examined average proportions of the parental alleles for genes belonging to recognition, signaling, and effector functional groups (Figure 4.5A). As with overall expression levels, we see that the patterns of allelic expression for these samples vary among the functional groups. While all samples show proportions of the *D. melanogaster* allele close to 0.5, the signaling genes show the lowest level of variability in these proportions. Signaling genes also appear to have lower variance in expression levels compared to genes in the effector group, yet this difference does not appear as striking. The signaling genes also show an effect of genotype on allelic proportions; the proportion of *D. melanogaster* allele in hybrids is significantly greater than the proportion of *D. melanogaster* transcript within the parental totals (ANOVA, $P = 0.00124$).

Genome-wide, we examined average proportions of the parental alleles in genes belonging to pathways as defined in KEGG. Table 4.2 lists those with significant effects of infection or a genotype by infection interaction on these proportions. Averages across samples are shown for those pathways where the genotype by infection effect has $P < 0.05$ (ANOVA), in Figure 4.5B. These pathways

Figure 4.5 Average proportions of *D. melanogaster* allele in expressed genes in (A) functional groups within the innate immune pathways and in (B) genes involved in non-immune pathways. Allelic proportions are shown for hybrid expression before and after infection (white and light gray bars), and proportions of *D. melanogaster* transcript out of total parental transcript before and after infection are shown (dark gray and black bars). (C) Average proportions of *D. melanogaster* allele in infected hybrids for all pathways tested. Dashed lines illustrate proportions equal to 0.5 on each plot. Asterisks indicate pathways with a significant genotype effect (in plot A) or a significant genotype x infection interaction effect (in plot B) on allelic proportions: * $P < 0.05$; ** $P < 0.01$. (Abbreviations for pathways: Aminosugars = Aminosugars metabolism; Glycophos. = Glycophospholipid metabolism; Purine = Purine metabolism; Glycan deg. = Glycan structures – degradation; SNARE = SNARE interactions in vesicular transport.)

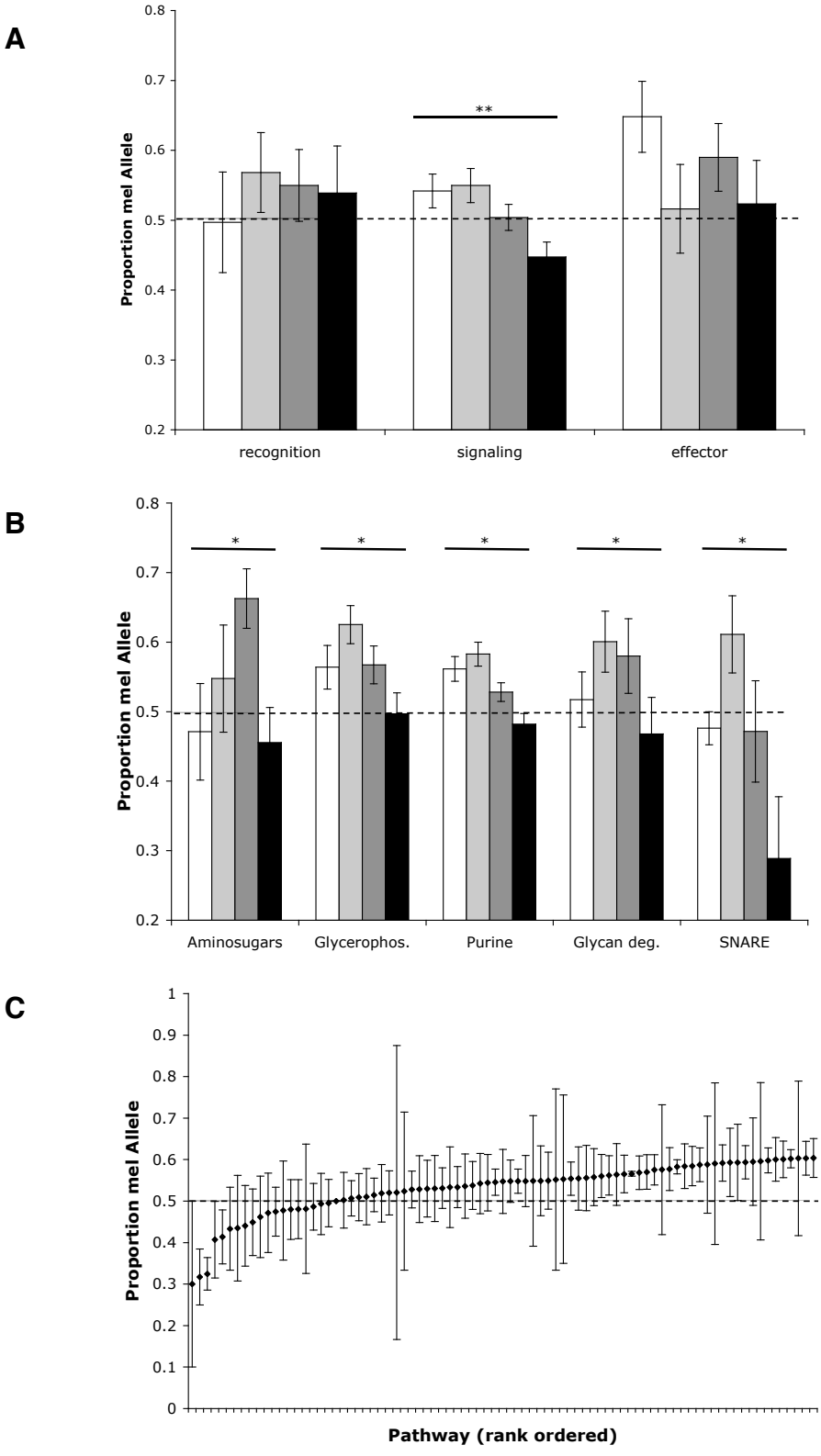

A
Proportion mel Allele
0.8
0.7
0.6
0.5
0.4
0.3
0.2
**
recognition
signaling
effector
B
Proportion mel Allele
0.8
0.7
0.6
0.5
0.4
0.3
0.2
*
*
*
*
*
Aminosugars
Glycerophos.
Purine
Glycan deg.
SNARE
C
Proportion mel Allele
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Pathway (rank ordered)

Table 4.2 Pathways Showing Significant Effects of Infection on Allelic Proportions

Pathway Name	Infection	Genotype x Infection
Aminosugars metabolism	0.294	**0.024**
Glycerophospholipid metabolism	0.878	**0.03**
Purine metabolism	0.444	**0.032**
Glycan structures - degradation	0.771	**0.041**
SNARE interactions in vesicular transport	0.716	**0.048**
Bile acid biosynthesis	**0.004**	0.506
Calcium signaling pathway	**0.018**	0.633
3-Chloroacrylic acid degradation	**0.021**	0.294
Apoptosis	**0.021**	0.503
TGF-beta signaling pathway	**0.035**	0.793
Lysine degradation	**0.048**	0.202

ANOVA *P* -values < 0.05 indicated in bold

show a variety of patterns for the proportions of the parental alleles, potentially indicative of heterogeneous levels of divergence in the regulation of expression of these genes after infection. Similarly, across all pathways tested, the proportions of the *D. melanogaster* allele among hybrid transcripts vary substantially; Figure 4.5C shows a plot of the proportion of *melanogaster* allelic expression among means for transcripts in each pathway in hybrids after infection.

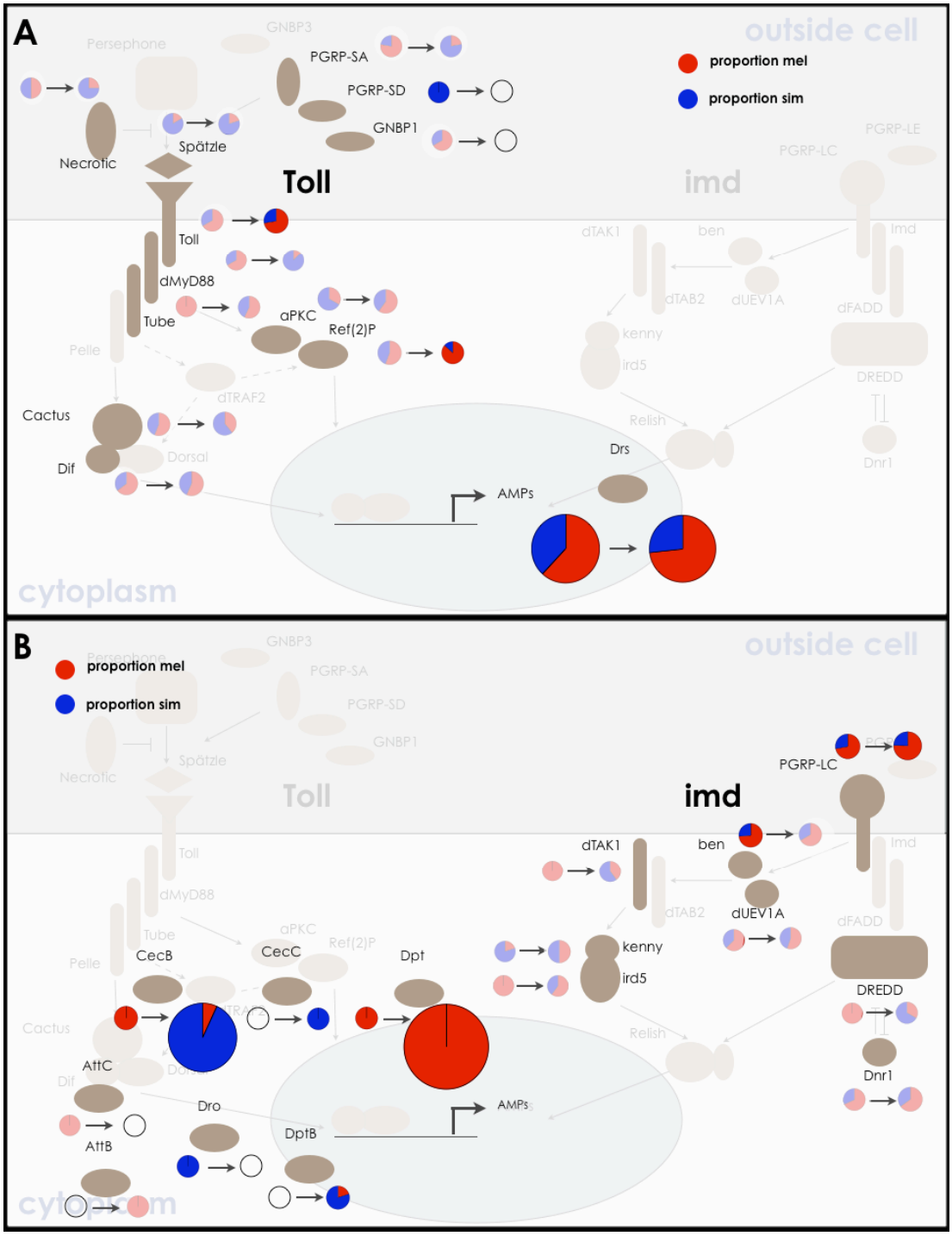
For a more specific evaluation of variability in allelic transcript abundance throughout the immune response, we examined allelic proportions in individual genes throughout the Toll and imd pathways (Figure 4.6). From this, we see some variations in allelic proportions throughout the network, particularly in the downstream components. On a gene-by-gene basis, however, the variability was sufficiently chaotic that we have little power to detect pathway-wide patterns.

Discussion

Using the power of an RNA-seq approach to quantify expression differences among samples, we were able to assay changes in transcript level in response to infection in Drosophila species and their F1 hybrids, as well as changes in relative transcript abundance of the two parental alleles in the hybrids. From this, we found further evidence for divergent regulation of the immune response between *D. melanogaster* and *D. simulans* as revealed through dysregulation of transcription after infection in their F1 hybrid, particularly in examinations of separate functional groups throughout the innate immune pathways. Additionally, we found effects of dysregulation of the response to infection, evinced by the irregular control of non-immune pathways genome-wide in hybrids.

With this experimental design, we have used internal comparisons between uninfected and infected samples to compare levels of transcript abundance across

Figure 4.6 Differences in proportions of the two parental alleles in hybrids within the transcripts of genes in the Toll (A) and imd (B) pathways before and after infection (to the left and right of each arrow, respectively). For each gene, proportion of the *D. melanogaster* allele is shown in red and the *D. simulans* allele is shown in blue; size of the circle corresponds to total level of transcript for that gene. Dark colors indicate a significant difference between the abundance of the two alleles (binomial test, $P < 0.05$). Open circles indicate absence of transcription. Pathway genes and interactions included based on information in previous studies (ARBOUZOVA and ZEIDLER 2006; FERRANDON *et al.* 2007; FOLEY and O'FARRELL 2004; LECLERC and REICHHART 2004; STRONACH and PERRIMON 2002; WASSARMAN *et al.* 1995).


A
outside cell
proportion mel
proportion sim
Persephone
GNBP3
PGRP-SA
PGRP-SD
GNBP1
Spätzle
Necrotic
Toll
imd
PGRP-LE
PGRP-LC
Toll
dMyD88
Tube
aPKC
Ref(2)P
Pelle
dTRAF2
Cactus
Dif
Dorsal
dTAK1
ben
dTAB2
dUEV1A
Imd
dFADD
kenny
ird5
DREDD
Relish
Dnr1
Drs
AMPs
cytoplasm
B
outside cell
proportion mel
proportion sim
GNBP3
PGRP-SA
PGRP-SD
GNBP1
Spätzle
Necrotic
Toll
imd
PGRP-LC
Toll
dMyD88
Tube
aPKC
Ref(2)P
Pelle
CecB
CecC
Dpt
dTAK1
ben
dTAB2
dUEV1A
Imd
dFADD
kenny
ird5
DREDD
Relish
Dnr1
Cactus
AttC
Dif
AttB
Dro
DptB
AMPs
cytoplasm

genes genome-wide. Along with normalizations for transcript length, this allows us to be fairly confident that any different patterns that we detect among genes should have a biological basis. Furthermore, this platform has allowed us more flexibility than microarrays could; we can compare transcript levels from diverged species and their hybrids without concern of probe divergence issues, and we are able to detect differences between the species at divergent sites throughout the transcripts, providing for quantification of allele-specific expression in the F1 hybrids.

As we observed in a previous study (Chapter 3), the three functionally distinct groups of immune-related genes (recognition, signaling and effector) display dissimilar patterns of hybrid dysregulation, indicative of different patterns of evolution operating on the members of these groups. More specifically, both of these experiments provide evidence that genes encoding effector proteins can be substantially dysregulated in hybrids, relative to parental expression levels, in response to bacterial infection as well as in uninfected flies. This consistent pattern indicates divergence in the regulation of the innate immune pathways manifesting in overactivation of the transcription of the downstream genes. Most strikingly, it appears that genes in the effector group are generally overexpressed in the face of hybrid breakdown of some regulatory mechanisms. This seems to indicate that the default state of the immune system is constitutive activation, while negative regulation to achieve appropriate levels of expression is not conserved enough between species to be maintained in the context of interspecific hybrids. This is further supported by the observation that immune-related effectors are over-expressed even in the absence of bacterial challenge.

While the members of the well-defined innate immune pathways may show the strongest responses to infection, a systemic bacterial invasion will likely lead to regulatory differences in multiple gene regulatory pathways throughout an organism.

By examining genes in non-immune pathways, we can explore greater impacts of the immune activation in the F1 hybrids. Using KEGG pathway annotations, we were able to group transcribed genes throughout the genome by their functional pathway classification and test those pathways for significant responses to infection in transcript differences. While any number of genes throughout the genome may show expression differences before and after infection, by testing pathways as groups, we may detect coordinated regulation of a pathway that is disrupted in the hybrids. With this, we do find numerous non-immune pathways that show significantly different responses to infection in hybrids and in the parental species, and interestingly, many of these are involved in metabolism-related functions. This could be indicative of a larger effect of dysregulation of the immune response in hybrids: since hybrids appear to over-express and/or over-induce numerous genes throughout the immune pathways while maintaining a normally functioning ability to clear bacteria after infection (Chapter 3), it stands to reason that they may suffer a cost and that other systems may be under-regulated, which is consistent with patterns of some non-immune dysregulation apparent here.

Beyond overall transcription levels in genes throughout the genome, we have also examined allele-specific transcript abundances in the hybrid flies. If proportions of the parental alleles stay the same after infection, both alleles may appear to be regulated at the same level. When we observe average allelic proportions for the immune genes, based on groups of separate immune-related functions, we again see distinct patterns among the groups. Both the recognition and effector groups, though, show relatively higher levels of variance than the signaling group of genes; this may indicate less coordinated control of expression of genes with recognition and effector function in hybrids. On average, though, most of the immune genes appear to have approximately equal proportions of the two allelic transcripts. Strikingly, many of the

non-immune pathways appear to have greater departure from equal proportions of the two parental transcripts in the hybrids, both before and after infection. Some of this may reflect divergent controls of these pathways in the parental species, yet a number of these pathways have significantly different levels of the alleles in the hybrids relative to the transcript levels in the parents, indicative of regulatory interaction in the hybrids.

By sequencing transcripts from hybrid and parental flies before and after bacterial infection, we are able to infer differences among these genotypes in the regulation of gene expression after immune challenge. It should be noted, though, that change in transcript abundance is only one aspect of the regulation of the immune response; this assay does not take into account control mechanisms such as rate of translation, post-translational modification, or protein stability. Nevertheless, changes in transcript levels have been observed previously for genes throughout the genome in response to infection (APIDIANAKIS *et al*. 2005; DE GREGORIO *et al*. 2001; IRVING *et al*. 2001), and these can be disrupted in immune pathway mutants (DE GREGORIO *et al*. 2002). Thus, while transcription may not be the only portion of the immune response that can vary among genotypes, it is apparent that it can serve as a proxy for gene regulation among flies and that differences in transcript level are indicative of distinct responses to infection.

In this study, we have uncovered further evidence for dysregulation of the hybrid response to infection, indicative of widespread regulatory divergence of genes throughout the genome between *D. melanogaster* and *D. simulans*. Overall, this emphasizes that in hybrid individuals, numerous parts of the parental genomes may interact to yield distinct patterns of expression, as a result of parental regulatory divergence. Furthermore, a perturbation such as immune response activation has the potential to affect regulation – and dysregulation – of genes throughout the genome.

REFERENCES

APIDIANAKIS, Y., M. N. MINDRINOS, W. XIAO, G. W. LAU, R. L. BALDINI *et al.*, 2005 Profiling early infection responses: *Pseudomonas aeruginosa* eludes host defenses by suppressing antimicrobial peptide gene expression. Proc Natl Acad Sci U S A **102:** 2573-2578.

AUGER, D. L., A. D. GRAY, T. S. REAM, A. KATO, E. H. COE, JR. *et al.*, 2005 Nonadditive gene expression in diploid and triploid hybrids of maize. Genetics **169:** 389-397.

BRAY, N., and L. PACHTER, 2004 MAVID: constrained ancestral alignment of multiple sequences. Genome Res **14:** 693-699.

CLOONAN, N., A. R. FORREST, G. KOLLE, B. B. GARDINER, G. J. FAULKNER *et al.*, 2008 Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods **5:** 613-619.

DE GREGORIO, E., P. T. SPELLMAN, G. M. RUBEN and B. LEMAITRE, 2001 Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. Proc Natl Acad Sci U S A **98:** 12590-12595.

DE GREGORIO, E., P. T. SPELLMAN, P. TZOU, G. M. RUBIN and B. LEMAITRE, 2002 The Toll and Imd pathways are the major regulators of the immune response in Drosophila. EMBO J **21:** 2568-2579.

DICKINSON, W. J., R. G. ROWAN and M. D. BRENNAN, 1984 Regulatory Gene Evolution: Adaptive Differences of Alcohol Dehydrogenase in *Drosophila melanogaster* and *Drosophila simulans*. J. Hered. **52:** 215-225.

HEGARTY, M. J., J. M. JONES, I. D. WILSON, G. L. BARKER, J. A. COGHILL *et al.*, 2005 Development of anonymous cDNA microarrays to study changes to the

Senecio floral transcriptome during hybrid speciation. Mol Ecol **14:** 2493-2510.

FERRANDON, D., J. L. IMLER, C. HETRU and J. A. HOFFMANN, 2007 The Drosophila systemic immune response: sensing and signaling during bacterial and fungal infections. Nat Rev Immunol **7:** 862-874.

FOLEY, E., and P. H. O'FARRELL, 2004 Functional dissection of an innate immune response by a genome-wide RNAi screen. PLoS Biol **2:** E203.

IRVING, P., L. TROXLER, T. S. HEUER, M. BELVIN, C. KOPCZYNSKI *et al.*, 2001 A genome-wide analysis of immune responses in Drosophila. Proc Natl Acad Sci U S A **98:** 15119-15124.

KANEHISA, M., and S. GOTO, 2000 KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res **28:** 27-30.

LANDRY, C. R., D. L. HARTL and J. M. RANZ, 2007 Genome clashes in hybrids: insights from gene expression. Heredity **99:** 483-493.

LAZZARO, B. P., T. B. SACKTON and A. G. CLARK, 2006 Genetic variation in *Drosophila melanogaster* resistance to infection: a comparison across bacteria. Genetics **174:** 1539-1554.

LAZZARO, B. P., B. K. SCEURMAN and A. G. CLARK, 2004 Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. Science **303:** 1873-1876.

LECLERC, V., and J. M. REICHHART, 2004 The immune response of *Drosophila melanogaster*. Immunol Rev **198:** 59-71.

LI, H., J. RUAN and R. DURBIN, 2008 Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res **18:** 1851-1858.

MICHALAK, P., and M. A. NOOR, 2003 Genome-wide patterns of expression in Drosophila pure species and hybrid males. Mol Biol Evol **20:** 1070-1076.

MOEHRING, A. J., K. C. TEETER and M. A. NOOR, 2007 Genome-wide patterns of expression in Drosophila pure species and hybrid males. II. Examination of multiple-species hybridizations, platforms, and life cycle stages. Mol Biol Evol **24:** 137-145.

MORTAZAVI, A., B. A. WILLIAMS, K. MCCUE, L. SCHAEFFER and B. WOLD, 2008 Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods **5:** 621-628.

NAGALAKSHMI, U., Z. WANG, K. WAERN, C. SHOU, D. RAHA *et al.*, 2008 The transcriptional landscape of the yeast genome defined by RNA sequencing. Science **320:** 1344-1349.

PARKER, H. R., D. P. PHILIPP and G. S. WHITT, 1985 Gene regulatory divergence among species estimated by altered developmental patterns in interspecific hybrids. Mol Biol Evol **2:** 217-250.

RANZ, J. M., K. NAMGYAL, G. GIBSON and D. L. HARTL, 2004 Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. Genome Res **14:** 373-379.

REILAND, J., and M. A. NOOR, 2002 Little qualitative RNA misexpression in sterile male F1 hybrids of *Drosophila pseudoobscura* and *D. persimilis*. BMC Evol Biol **2:** 16.

RENAULT, S., A. W. NOLTE and L. BERNATCHEZ, 2009 Gene Expression Divergence and Hybrid Misexpression between Lake Whitefish Species Pairs (*Coregonus* spp. Salmonidae). Mol Biol Evol **26:** 925-936.

SACKTON, T. B., B. P. LAZZARO, T. A. SCHLENKE, J. D. EVANS, D. HULTMARK *et al.*, 2007 Dynamic evolution of the innate immune system in Drosophila. Nat Genet **39:** 1461-1468.

SWANSON-WAGNER, R. A., Y. JIA, R. DECOOK, L. A. BORSUK, D. NETTLETON *et al.*, 2006 All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci U S A **103:** 6805-6810.

TRUE, J. R., and E. S. HAAG, 2001 Developmental system drift and flexibility in evolutionary trajectories. Evol Dev **3:** 109-119.

WHITT, G. S., D. P. PHILIPP and W. F. CHILDERS, 1977 Aberrant Gene Expression during the Development of Hybrid Sunfishes (Perciformes, Teleostei). Differentiation **9:** 97-109.

CHAPTER 5

DISRUPTION OF THE INNATE IMMUNE RESPONSE IN MUTATION-PERTURBED INTERSPECIFIC DROSOPHILA HYBRIDS

Abstract

The pathways involved in the innate immune response in Drosophila make up a complex regulatory network, including numerous proteins that are involved in multiple interactions in response to infection. Genes encoding members of these pathways have been shown to bear different levels of variation and distinct patterns of divergence between species, indicative of unique selective pressures acting on genes performing separate functions within the immune response. Previously, we found evidence for dissimilar regulatory evolution among different functional groups of immune genes by quantifying expression differences in *D. melanogaster*, *D. simulans*, and their F1 hybrids before and after infection, introducing divergent regulatory mechanisms throughout the immune pathways (Chapter 3; Chapter 4). To further inspect the sensitivity of the system to divergent genetic interactions in the innate immune pathways, we examine immune responses of hybrids bearing mutant alleles for one copy of certain pathway components to determine the effects of interactions between the remaining allele and diverged interacting proteins on immune function. With these, we were able to pinpoint the Dredd protein as one that has evolved such that the disturbance of its conspecific interactions can lead to decreased function of the immune response.

Introduction

Drosophila has become one of the pre-eminent models for studying the innate immune system, not only because of the vast array of techniques and resources

available for these flies, but also due to the high degree of homology present between their humoral immune pathways and those in other species. Numerous genetic studies in Drosophila have clarified the members of and interactions within the innate immune pathways. Furthermore, comparative analyses among Drosophila species have facilitated evolutionary queries into the system. Many of these studies have focused on sequence diversity of immunity genes within and between species (CLARK and WANG 1997; DATE *et al.* 1998; LAZZARO and CLARK 2003; RAMOS-ONSINS and AGUADE 1998; SACKTON *et al.* 2007; SCHLENKE and BEGUN 2003), while others have examined variation occurring in natural populations of *D. melanogaster* (LAZZARO *et al.* 2006; LAZZARO *et al.* 2004; SACKTON *et al*. *submitted*). One of the key trends from these studies is the disparity in levels of variation and natural selection apparent among genes involved in different functions within the immune pathways.

More recently, we have examined the impacts of interspecific hybridization on immune function, in order to uncover diverged portions of the immune networks between *D. melanogaster* and *D. simulans* (Chapter 3; Chapter 4). With these studies, we found differences in hybrid dysregulation of the expression of immune genes among different functional groups, consistent with previous findings of distinct patterns of evolution for these groups. Furthermore, we found a preponderance of over-expressed genes among the downstream effectors of the pathways, providing evidence for a propagation of dysregulation throughout these pathways in F1 hybrids. These investigations, similar to those that have previously assayed expression or phenotype differences in F1 hybrid individuals (DICKINSON *et al.* 1984; MOEHRING *et al.* 2007; PARKER *et al.* 1985; RANZ *et al.* 2004; WHITT *et al.* 1977), revealed systemic disruptions in hybrid phenotypes, relative to those of the parental species, due to the juxtaposition of two diverged genomes within an individual.

While these studies are useful for examining the functional divergence of one or more networked pathways, they can only uncover dysfunctional phenotypes when the interactions between one or more diverged gene pairs have a dominant effect. More specifically, if two genes whose products interact have evolved separately through distinct lineages, they could potentially yield less fit phenotypes when the diverged gene products interact in the interspecific hybrids – in some cases leading to hybrid sterility or inviability and ultimately, speciation (BATESON 1909; DOBZHANSKY 1934; MULLER 1942). Such phenotypes will only manifest in the hybrid organisms, though, if they are dominant to the conspecific interactions – since F1 hybrids have both parental alleles for each component of an interaction, normal complexes may still form in these individuals.

To further probe the divergence in the components of the immune system between species of Drosophila, we have introduced mutant *D. melanogaster* alleles into F1 hybrid backgrounds. For this, we mated *D. melanogaster* females bearing mutations in one of several genes throughout the Toll and imd pathways to *D. simulans* males. These crosses allowed us to assay immune response phenotypes in flies bearing genotypes perturbed beyond interspecific hybridization. By including a mutant *D. melanogaster* allele, we expect that any evolved interactions involving that gene's product will be placed under greater strain; if interspecific interactions are less fit, the effect of a mutant allele should be greater in a hybrid background than in a *D. melanogaster* genetic background. Since some genes in the immune networks, particularly *Relish*, *Dredd*, *ird5*, *key*, and *Dnr1* in the imd pathway, have been shown previously to display evidence of positive selection among species in the melanogaster subgroup (SACKTON *et al.* 2007), we expect that some of these may have evolved between *D. melanogaster* and *D. simulans* to such a degree that their interactions within the immune response may be disrupted. Using this approach, we have

identified genes in the humoral pathways displaying evidence of divergence between the two species, leading to disruption of normal immune function in these mutant hybrids.

Materials and Methods

Fly Lines and Crosses: To evaluate the impact of targeted disruptions of the innate immune response in the context of a hybrid genetic background, we mated *D. melanogaster* females bearing mutations in genes throughout the immune pathways to *D. simulans* males (strain MD199). We used mutant stocks containing P element insertions disrupting genes in the Toll and imd pathways (Exelixis Collection, Harvard Medical School and Bloomington Drosophila Stock Center at Indiana University; (THIBAULT *et al.* 2004)). The stocks we used here included mutants for insertions in 13 genes, listed in Table 5.1 (see Table C.1 for a detailed stock information).

These mutant stocks were created in one of two laboratory strains of *D. melanogaster* – either w^{1118} or $y^{1}w^{67c23}$. To control for background effects of these genotypes, these two "wild type" strains were also included in the crossing schemes, to yield wild type *D. melanogaster* alleles for the genes of interest, in the appropriate genetic backgrounds. Thus, we mated virgin females from each *D. melanogaster* line to males of the "wild type" *D. melanogaster* lines, as well as to *D. simulans* males, with approximately 10 males and 10 females for each cross vial (Figure 5.1). We also included crosses of *D. simulans* females and males, to represent the second parental genotype. Out of the 13 mutant stocks of *D. melanogaster*, 10 were able to hybridize with *D. simulans*, so these were the only ones used to assay immune response phenotypes, as listed in Table 5.1.

Bacterial Cultures and Infections: To assay immune response phenotypes among the different genotypes, we infected the flies with either Gram-negative

Table 5.1 Genes disrupted in Exelixis P element insertion lines

Pathway	Gene	Location of Insert	% AA Identity mel vs. sim	Successful hybridization?
Toll	*spz*	coding	91	no
	Myd88	5' UTR	77	YES
	tub	5' UTR	64	no
	pll	3' UTR	94	YES
	cact	5' UTR	77	YES
	Dif	5' UTR	98	no
imd	*PGRP-LC*	3' intergenic	84	YES
	imd	5' intergenic	97	YES
	BG4	coding	76	YES
	Dredd	5' UTR	85	YES
	ird5	coding	87	YES
	key	coding	81	YES
	Rel	5' UTR	85	YES

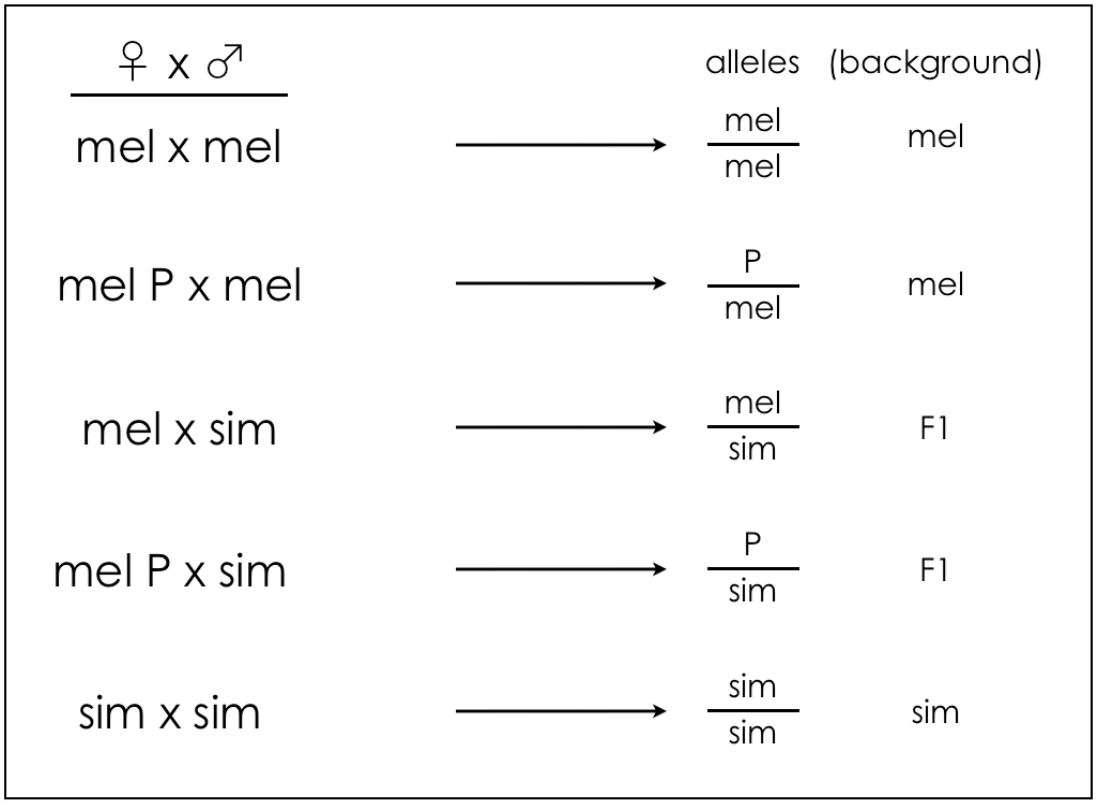


Figure 5.1 Cross scheme used for mutant hybrids and control crosses. "P" indicates *melanogaster* allele with P element insertion.

Serratia marcescens or Gram-positive *Enterococcus faecalis*. These bacteria were chosen because of their use in previous studies of immune response in Drosophila (LAZZARO *et al.* 2004). The *S. marcescens* stock was derived from ATCC strain 13880, and the *E. faecalis* was derived from that used by (LAZZARO *et al.* 2006) (identified via 16S rDNA sequence and results of API 20Strep substrate utilization testing). Cultures were grown overnight from freezer stocks to $OD_{600} \sim 1.0$. To infect the flies, we pricked their thoraces with 0.1-mm tungsten needles (Fine Science Tools, Foster City, CA) that had been dipped in the bacterial cultures.

Bacterial Load Quantification: To quantify microbial clearing ability, we measured bacterial load in the flies at 12 and 25 hours after infection with *S. marcescens* or *E. faecalis*. For this, we combined females in groups of three in 500µl of LB broth and homogenized the flies. Up to 10 replicates of each genotype were infected each day, on four separate days; in total, between 13 and 34 replicates were infected for each. We then plated the homogenates onto LB plates using a spiral plater (Spiral Biotech, Bethesda, MD) and allowed colonies on the plates to grow overnight. We counted these using a colony counter, giving us an estimate of bacterial load, or colony-forming units per ml (cfu/ml). We also inspected plates visually to ensure that the colonies counted were consistent with the size and morphology expected.

Quantification of Transcript Abundance using Pyrosequencing: To evaluate differences in transcript abundance of the antimicrobial peptide *Diptericin* (*Dpt*) as a result of differential regulation of the immune response among these genotypes, we used Pyrosequencing (Qiagen, Valencia, CA). This allowed us to assay relative transcript levels between two parental alleles (in the interspecific hybrids), as well as quantify relative levels of total transcript between samples from uninfected and infected flies. For this, we froze flies in liquid nitrogen 12 hours after infection with *S. marcescens* or *E. faecalis*, or without infection, in three replicate pools of

approximately 13-17 flies each. For each sample, we isolated mRNA using a Trizol:chloroform extraction, and then we synthesized cDNA using oligo-dT primers.

For the Pyrosequencing, we amplified cDNA from each sample using primers for *Dpt* (see Table C.2 for primer and probe sequences). We treated the resulting products according to Pyrosequencing protocols in order to sequence the region of interest, including one or more sites with divergent bases between the two parental species. Overall transcript level for each genotype was determined by averaging the heights of the peaks across the non-divergent bases as well as across biological replicates. In order to estimate differences in transcript abundance between the two alleles in the interspecific hybrids, we normalized the peak heights of both alleles by the parental levels (to correct for background effects of one or both alleles).

Results

Effect of Genotype on Bacterial Clearing Ability: To evaluate the phenotypic effects of divergent genes interacting within the innate immune pathways, we quantified differences in immunocompetence of immune pathway mutants in *D. melanogaster* and interspecific hybrid backgrounds. Figure 5.2 shows the bacterial load differences between the mutant and wild type alleles in each background for each of the mutants after infection with *S. marcescens*. Two of the mutants, *Dredd* and *Myd88*, show nearly significant effects of an allele x background interaction on load (ANOVA, $P = 0.079$, 0.064, respectively). More specifically, one copy of the mutant *Dredd* allele associates with an increased bacterial load after Gram-negative infection – only in the context of a hybrid genetic background. Conversely, the presence of a copy of the *Myd88* mutant allele appears to lead to an increase in bacterial load only in the *D. melanogaster* background, possibly indicating a specificity of the effect between the genes belonging to Toll and imd pathways in response to this infection.

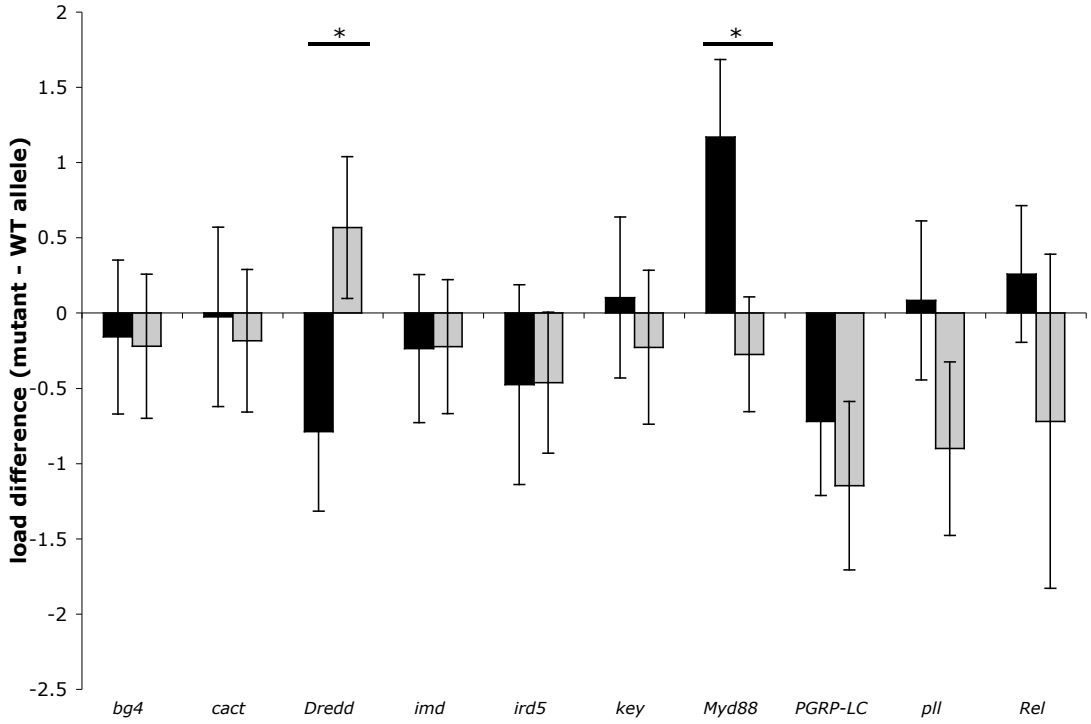


Figure 5.2 Differences in bacterial load after infection with *S. marcescens* between flies bearing mutant and wild type alleles. Differences between these genotypes is shown for flies with *D. melanogaster* (black bars) and F1 hybrid (gray bars) genetic backgrounds. Asterisks indicate an allele by background interaction effect with $P < 0.1$.

To examine the effect of infection type (Gram-positive vs. Gram-negative) on the sensitivity to perturbation of members of Toll and imd pathways, we quantified bacterial load after infection with *E. faecalis* in crosses involving *Dredd* and *Myd88* mutant alleles. The bacterial clearing phenotypes of these flies are compared to those after infection with *S. marcescens* in Figure 5.3. With these infections, however, neither mutant allele appears to have an effect on the load phenotype that is dependent upon the genetic background.

Effect of Genotype on Transcription Levels: As a proxy for the induction of an immune response, we quantified transcript levels of *Diptericin* (*Dpt*), an antimicrobial peptide (AMP) induced upon activation of the imd pathway, shown in Figure 5.4. We find that *Dpt* expression depends upon fly genotype, particularly after *S. marcescens* infection. The presence of a mutant *Dredd* allele decreases the level of induction of *Dpt* after infection with *S. marcescens*. Most notably, though, when the *Dredd* mutant allele occurs in an F1 hybrid background, rather than in a *D. melanogaster* background, the level of *Dpt* transcript actually shows a significant decrease in response to Gram-negative infection (*t*-test, $P = 2.56 \times 10^{-5}$), again indicating a detrimental interaction between disruption of the *D. melanogaster* allele and a hybrid genetic background, in terms of maintaining a proper immune response.

This effect of this interaction on *Dpt* transcription appears specific for the response to infection with *S. marcescens* – suppression of transcription after *E. faecalis* infections does not appear to occur in the presence of mutant alleles. Additionally, the effect of a mutant allele in a hybrid background on *Dpt* induction appears specific to disruptions of *Dredd* and not *Myd88*. These results are not surprising, though, given the fact that Gram-negative bacteria should elicit response via the imd pathway (including *Dredd*) more than the Toll pathway (including

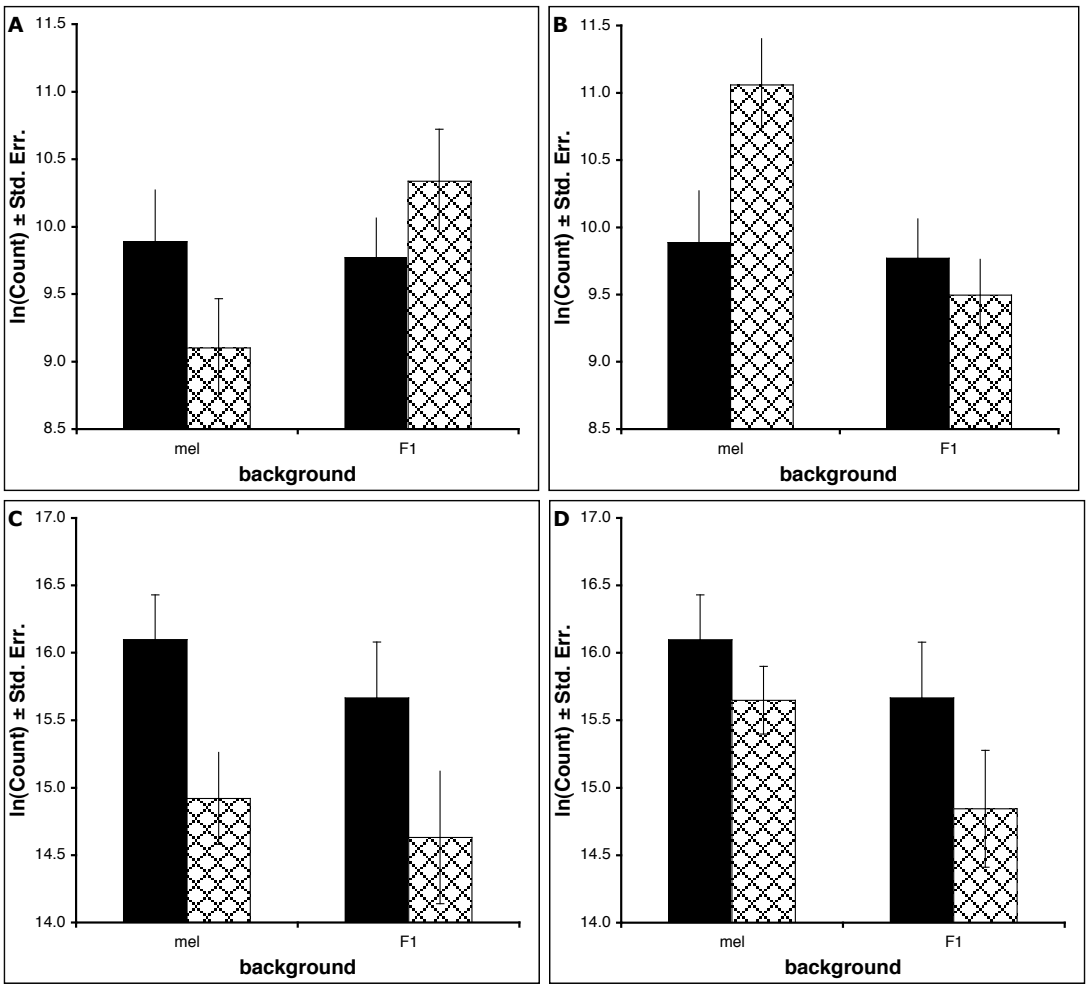


Figure 5.3 Bacterial load after infection in flies bearing wild type alleles (black bars) and mutant alleles (hatched bars) for *Dredd* (A & C) and *Myd88* (B & D) mutants. Bacterial load shown for flies after infection with either Gram-negative *S. marcescens* (A & B) or Gram-positive *E. faecalis* (C & D).

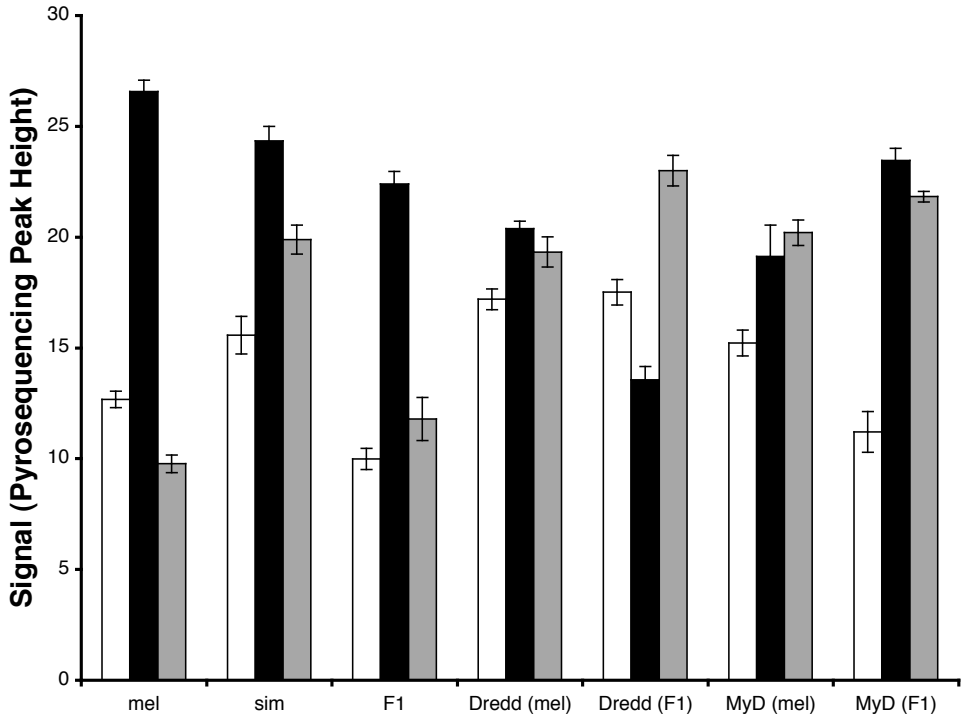


Figure 5.4 Transcript levels of *Dpt* in uninfected flies (white bars), and in flies infected with *S. marcescens* (black bars) or *E. faecalis* (gray bars). Genotype in parenthesis indicates genetic background for each mutant.

Myd88), and Gram-positive infection would not be expected to yield these same effects on *Dpt* expression through the imd pathway.

Dredd Genotype Has No Effect on Relative Transcript Levels of Parental Alleles: In addition to examining total levels of *Dpt* transcript in these flies, we also quantified relative abundance of the two parental alleles in the F1 hybrids. We found that hybrids have approximately equal abundances of the transcripts from each parental allele, and that the relative transcript levels do not differ significantly between either infection states or genotypes (ANOVA, $P = 0.585$, 0.7748, respectively). Figure 5.5 shows the average proportions of the *D. melanogaster* transcript for hybrids with or without *S. marcescens* infection and with wild type or mutant *D. melanogaster* alleles of *Dredd*.

Discussion

Previous studies have shown that interspecific hybrids can manifest disrupted phenotypes on a variety of scales – from systemic failures leading to inviability to more minor gene dysregulations, yielding expression patterns for genes throughout the genome that can deviate from those in the parental species. By examining dysfunctional phenotypes in F1 hybrids, divergence between the two parental species in the regulation of those phenotypes can be inferred. Such studies have revealed delayed development and dysregulated expression of enzymes during development (PARKER *et al.* 1985), expression levels outside those of parents for many transcripts genome-wide (RANZ *et al.* 2004), and dysregulation of the immune system in response to infection (Chapter 3; Chapter 4), in interspecific F1 hybrids relative to their parental species. While a number of these queries have focused on targeted portions of the hybrid physiology, these patterns of dysregulation could be influenced by divergence in numerous genes throughout the system of interest.

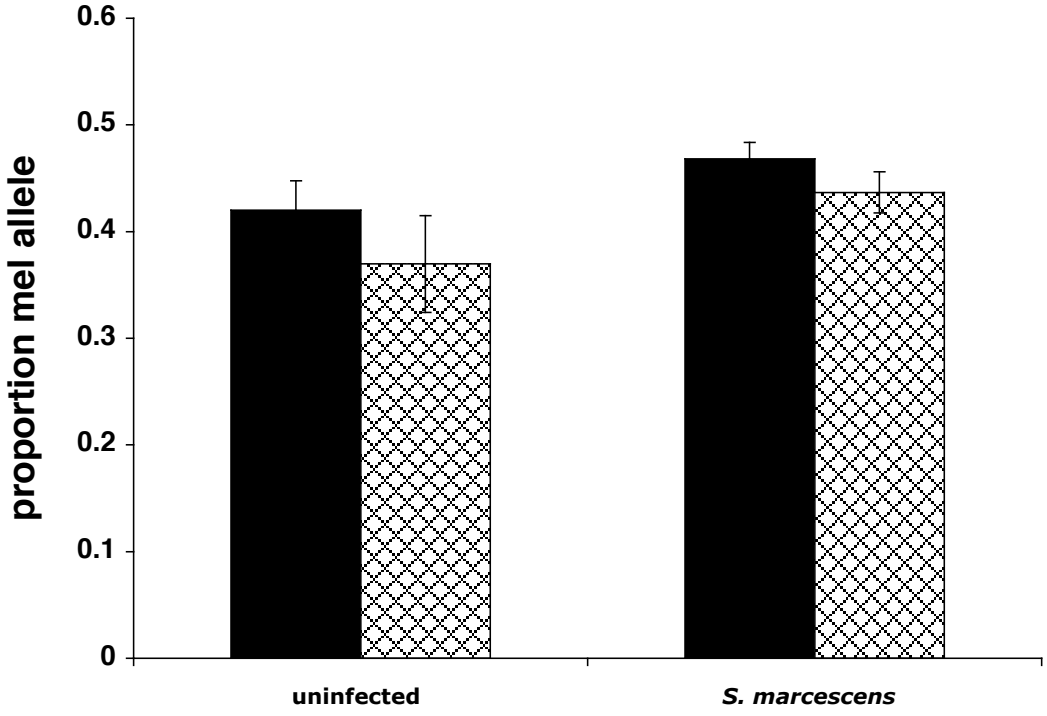


Figure 5.5 Proportion of *Dpt* transcript from the *D. melanogaster* parental allele in F1 hybrids before and after infection. Flies with wild type (black bars) and *Dredd* mutant (hatched bars) alleles are shown.

Here we use a more targeted approach and have found evidence for interspecific divergence of one or more of the interactions in the immune pathways that involve the Dredd protein. This divergence does not disrupt immune response phenotypes in the F1 hybrids bearing wild type alleles, but when a mutant allele occurs in a hybrid background, the flies display signs of decreased immunocompetence. This is the expected result if interactions involving Dredd have evolved in *D. melanogaster* and *D. simulans* such that disrupting one parental allele limits the hybrid's ability to form functional interactions. In this case, the interspecific interactions are presumably less successful in the activation of the immune response, and thus the mutant allele has a greater effect in a hybrid than in a *D. melanogaster* genetic background. These results are consistent with the previous finding that *Dredd*, along with other members of the Relish cleavage complex, appear to evolving under positive selection (SACKTON *et al.* 2007).

Dredd was originally identified as a member of the caspase gene family, with a role during apoptosis in *D. melanogaster* (CHEN *et al.* 1998), and it was later found to be a crucial member of the imd pathway, necessary for response to Gram-negative bacterial infection in Drosophila (LEULIER *et al.* 2000). In such a role, there are numerous potential interactions involving Dredd that could be disrupted, leading to a dysfunctional immune response. Dredd exists downstream of BG4 in the pathway, and there is evidence that the two proteins bind via their death-inducing domains (HU and YANG 2000). Furthermore, Dredd was shown to be necessary for cleavage of Relish, an NF-κB protein, leading to the activation of AMP transcription (STOVEN *et al.* 2000), and Dredd and Relish have been shown to physically interact with one another (STOVEN *et al.* 2003). Additionally, negative regulators of the imd pathway, Dnr1 and Caspar, have been shown to repress the response to Gram-negative infection at the level of Dredd-mediated cleavage of Relish, with evidence that Dnr1 may bind

directly to Dredd (GUNTERMANN *et al.* 2009; KIM *et al.* 2006). One or more of these interactions could be disrupted in interspecific protein combinations and could therefore yield the phenotypes observed here.

The complexity of the pathways involved in the Drosophila innate immune response have the potential to include redundant mechanisms and therefore provide some degree of robustness to perturbations. On the other hand, in genetic networks, there may be one or more elements whose stability is critical for proper function of the network as a whole. By examining immune response in interspecific hybrids bearing mutant alleles for genes in the innate immune pathways, we have identified at least one instance of the latter, where interspecific combinations of alleles have disrupted the proper function of Dredd, leading to a compromised immune response. This highlights not only the critical nature of Dredd function within the pathways, but also the specificity of its interactions, which have diverged between Drosophila species.

REFERENCES

BATESON, W., 1909 Heredity and Variation in Modern Lights in *Darwin and Modern Science*, edited by A. C. SEWARD. Cambridge: at the University Press, New York.

CHEN, P., A. RODRIGUEZ, R. ERSKINE, T. THACH and J. M. ABRAMS, 1998 Dredd, a novel effector of the apoptosis activators reaper, grim, and hid in Drosophila. Dev Biol **201:** 202-216.

CLARK, A. G., and L. WANG, 1997 Molecular population genetics of Drosophila immune system genes. Genetics **147:** 713-724.

DATE, A., Y. SATTA, N. TAKAHATA and S. I. CHIGUSA, 1998 Evolutionary history and mechanism of the Drosophila *cecropin* gene family. Immunogenetics **47:** 417-429.

DICKINSON, W. J., R. G. ROWAN and M. D. BRENNAN, 1984 Regulatory Gene Evolution: Adaptive Differences of Alcohol Dehydrogenase in *Drosophila melanogaster* and *Drosophila simulans*. J. Hered. **52:** 215-225.

DOBZHANSKY, T., 1934 Studies on Hybrid Sterility I. Spermatogenesis in Pure and Hybrid *Drosophila pseudoobscura*. Cell and Tissue Research **21:** 169-223.

GUNTERMANN, S., D. A. PRIMROSE and E. FOLEY, 2009 Dnr1-dependent regulation of the Drosophila immune deficiency signaling pathway. Dev Comp Immunol **33:** 127-134.

HU, S., and X. YANG, 2000 dFADD, a novel death domain-containing adapter protein for the Drosophila caspase DREDD. J Biol Chem **275:** 30761-30764.

KIM, M., J. H. LEE, S. Y. LEE, E. KIM and J. CHUNG, 2006 Caspar, a suppressor of antibacterial immunity in Drosophila. Proc Natl Acad Sci U S A **103:** 16358-16363.

LAZZARO, B. P., and A. G. CLARK, 2003 Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. Mol Biol Evol **20:** 914-923.

LAZZARO, B. P., T. B. SACKTON and A. G. CLARK, 2006 Genetic variation in *Drosophila melanogaster* resistance to infection: a comparison across bacteria. Genetics **174:** 1539-1554.

LAZZARO, B. P., B. K. SCEURMAN and A. G. CLARK, 2004 Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. Science **303:** 1873-1876.

LEULIER, F., A. RODRIGUEZ, R. S. KHUSH, J. M. ABRAMS and B. LEMAITRE, 2000 The Drosophila caspase Dredd is required to resist gram-negative bacterial infection. EMBO Rep **1:** 353-358.

MOEHRING, A. J., K. C. TEETER and M. A. NOOR, 2007 Genome-wide patterns of expression in Drosophila pure species and hybrid males. II. Examination of multiple-species hybridizations, platforms, and life cycle stages. Mol Biol Evol **24:** 137-145.

MULLER, H. J., 1942 Isolating Mechanisms, Evolution and Temperature. Biol. Symp. **6:** 71-125.

PARKER, H. R., D. P. PHILIPP and G. S. WHITT, 1985 Gene regulatory divergence among species estimated by altered developmental patterns in interspecific hybrids. Mol Biol Evol **2:** 217-250.

RAMOS-ONSINS, S., and M. AGUADE, 1998 Molecular evolution of the *Cecropin* multigene family in Drosophila. functional genes vs. pseudogenes. Genetics **150:** 157-171.

RANZ, J. M., K. NAMGYAL, G. GIBSON and D. L. HARTL, 2004 Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. Genome Res **14:** 373-379.

SACKTON, T. B., B. P. LAZZARO, T. A. SCHLENKE, J. D. EVANS, D. HULTMARK *et al.*, 2007 Dynamic evolution of the innate immune system in Drosophila. Nat Genet **39:** 1461-1468.

SCHLENKE, T. A., and D. J. BEGUN, 2003 Natural selection drives Drosophila immune system evolution. Genetics **164:** 1471-1480.

STOVEN, S., I. ANDO, L. KADALAYIL, Y. ENGSTROM and D. HULTMARK, 2000 Activation of the Drosophila NF-kappaB factor Relish by rapid endoproteolytic cleavage. EMBO Rep **1:** 347-352.

STOVEN, S., N. SILVERMAN, A. JUNELL, M. HEDENGREN-OLCOTT, D. ERTURK *et al.*, 2003 Caspase-mediated processing of the Drosophila NF-kappaB factor Relish. Proc Natl Acad Sci U S A **100:** 5991-5996.

THIBAULT, S. T., M. A. SINGER, W. Y. MIYAZAKI, B. MILASH, N. A. DOMPE *et al.*, 2004 A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. Nat Genet **36:** 283-287.

WHITT, G. S., D. P. PHILIPP and W. F. CHILDERS, 1977 Aberrant Gene Expression during the Development of Hybrid Sunfishes (Perciformes, Teleostei). Differentiation **9:** 97-109.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Table A.1 Autosomal Genotyping Results for Restriction Fragment Length Polymorphisms in Genes Across Chromosomes 2 and 3

Gene / Line	Dro,AttAB	imd	Tehao	cact	PGRP-LC	DrsL	BG4	AttD
proportion consistent	0.9875	0.98125	1	1	1	1	1	1
FM7a		uncut	cut	cut	cut	uncut		cut
X2	cut	uncut	cut	cut	cut	uncut		cut
X3	cut	uncut	cut	cut	cut	uncut	cut	cut
X4	cut	uncut	cut	cut	cut	uncut	cut	cut
X6	cut	uncut	cut	cut	cut	uncut	cut	cut
X7	uncut	uncut	cut	cut	cut	uncut	cut	cut
X9	cut	uncut	cut	cut	cut	uncut	cut	cut
X10	cut	uncut	cut	cut	cut	uncut	cut	cut
X11	cut	uncut	cut	cut	cut	uncut	cut	cut
X12	cut	uncut	cut	cut	cut	uncut	cut	cut
X13	cut	uncut	cut	cut	cut	uncut	cut	cut
X14	cut	uncut	cut	cut	cut	uncut	cut	cut
X15	cut	uncut	cut	cut	cut	uncut	cut	cut
X16	cut	uncut	cut	cut	cut	uncut	cut	cut
X17	cut	uncut	cut	cut	cut	uncut	cut	cut
X22	cut	uncut	cut	cut	cut	uncut	cut	cut
X23	cut	uncut	cut	cut	cut	uncut	cut	cut
X24	cut	uncut	cut	cut	cut	uncut	cut	cut
X25	cut	uncut	cut	cut	cut	uncut	cut	cut
X26	cut	uncut	cut	cut	cut	uncut	cut	cut
X27	cut	uncut	cut	cut	cut	uncut	cut	cut
X28	cut	uncut	cut	cut	cut	uncut	cut	cut
X29	cut	uncut	cut	cut	cut	uncut	cut	cut
X31	cut	uncut	cut	cut	cut	uncut	cut	cut
X33	cut	uncut	cut	cut	cut	uncut	cut	cut
X34	cut	uncut	cut	cut	cut	uncut	cut	cut
X35	cut	uncut	cut	cut	cut	uncut	cut	cut
X36	cut	uncut	cut	cut	cut	uncut	cut	cut
X37	cut	uncut	cut	cut	cut	uncut	cut	cut
X38	cut	uncut	cut	cut	cut	uncut	cut	cut
X39	cut	uncut	cut	cut	cut	uncut	cut	cut
X40	cut	uncut	cut	cut	cut	uncut	cut	cut
X41	cut	uncut	cut	cut	cut	uncut		cut
X42	cut	uncut	cut	cut	cut	uncut	cut	cut
X43	cut	uncut	cut	cut	cut	uncut		cut
X44	cut	uncut	cut	cut	cut	uncut	cut	cut
X46	cut	uncut	cut	cut	cut	uncut	cut	cut
X47	cut	uncut	cut	cut	cut	uncut	cut	cut
X48	cut	uncut	cut	cut	cut	uncut	cut	cut
X49	cut	uncut	cut	cut	cut	uncut		cut
X50	cut	uncut	cut	cut	cut	uncut		cut
X51	cut	uncut	cut	cut	cut	uncut		cut
X52	cut	uncut	cut	cut	cut	uncut	cut	cut
X53	cut	uncut	cut	cut	cut	uncut	cut	cut
X54	cut	uncut	cut	cut	cut	uncut	cut	cut
X55	cut	uncut	cut	cut	cut	uncut	cut	cut
X56	cut	uncut	cut	cut	cut	uncut	cut	cut
X57	cut	uncut	cut	cut	cut	uncut	cut	cut
X58	cut	uncut	cut	cut	cut	uncut	cut	cut
X59	cut	uncut	cut	cut	cut	uncut	cut	cut
X60	cut	uncut	cut	cut	cut	uncut	cut	cut
X61	cut	cut	cut	cut	cut	uncut	cut	cut
X62	cut	uncut	cut	cut	cut	uncut	cut	cut
X63	cut	uncut	cut	cut	cut	uncut	cut	cut
X64	cut	uncut	cut	cut	cut	uncut	cut	cut
X65	cut	uncut	cut	cut	cut	uncut	cut	cut
X68	cut	uncut	cut	cut	cut	uncut	cut	cut
X69	cut	cut	cut	cut	cut	uncut	cut	cut

Table A.1 (continued)

X70	cut	uncut	cut	cut	cut	uncut	cut	cut
X71	cut	uncut	cut	cut	cut	uncut	cut	cut
X73	cut	cut	cut	cut	cut	uncut	cut	cut
X74	cut	uncut	cut	cut	cut	uncut	cut	cut
X75	cut	uncut	cut	cut	cut	uncut	cut	cut
X76	cut	uncut	cut	cut	cut	uncut	cut	cut
X79	cut	uncut	cut	cut	cut	uncut	cut	cut
X80	cut	uncut	cut	cut	cut	uncut	cut	cut
X81	cut	uncut	cut	cut	cut	uncut		cut
X83	cut	uncut	cut	cut	cut	uncut		cut
X84	cut	uncut	cut	cut	cut	uncut	cut	cut
X86	cut	uncut	cut	cut	cut	uncut		cut
X87	cut	uncut	cut	cut	cut	uncut	cut	cut
X88	cut	uncut	cut	cut	cut	uncut		cut
X89					cut			
X90	cut	uncut	cut	cut	cut	uncut		cut
X91	cut	uncut	cut	cut	cut	uncut	cut	cut
X92	cut	uncut	cut	cut	cut	uncut	cut	cut
X93	cut	uncut	cut	cut	cut	uncut	cut	cut
X94	cut	uncut	cut	cut	cut	uncut	cut	cut
X95	cut	uncut	cut	cut	cut	uncut	cut	cut
X96	cut	uncut	cut	cut	cut	uncut	cut	cut
X97	cut	uncut	cut	cut	cut	uncut		cut
X98	cut	uncut	cut	cut	cut	uncut	cut	cut
X99	cut	uncut	cut	cut	cut	uncut	cut	cut
X101	cut	uncut	cut	cut	cut	uncut		cut
X102	cut	uncut	cut	cut	cut	uncut		cut
X103	cut	uncut	cut	cut	cut	uncut	cut	cut
X104	other	uncut	cut	cut	cut	uncut	cut	cut
X105	cut	uncut	cut	cut	cut	uncut	cut	cut
X106	cut	uncut	cut	cut	cut	uncut	cut	cut
X107	cut	uncut	cut	cut	cut	uncut		cut
X108	cut	uncut	cut	cut	cut	uncut	cut	cut
X109	cut	uncut	cut	cut	cut	uncut	cut	cut
X110	cut	uncut	cut	cut	cut	uncut	cut	cut
X111	cut	uncut	cut	cut	cut	uncut	cut	cut
X112	cut	uncut	cut	cut	cut	uncut		cut
X113	cut	uncut	cut	cut	cut	uncut		cut
X114	cut	uncut	cut	cut	cut	uncut		cut
X115	cut	uncut	cut	cut	cut	uncut		cut
X116	cut	uncut	cut	cut	cut	uncut		cut
X117	cut	uncut	cut	cut	cut	uncut		cut
X119	cut	uncut	cut	cut	cut	uncut		cut
X122	cut	uncut	cut	cut	cut	uncut		cut
X123	cut	uncut	cut	cut	cut	uncut		cut
X125	cut	uncut	cut	cut	cut	uncut		cut
X126	cut	uncut	cut	cut	cut	uncut		cut
X127	cut	uncut	cut	cut	cut	uncut		cut
X128	cut	uncut	cut	cut	cut	uncut		cut
X130	cut	uncut	cut	cut	cut	uncut		cut
X131	cut	uncut	cut	cut	cut	uncut		cut
X134	cut	uncut	cut	cut	cut	uncut		cut
X136	cut	uncut	cut	cut	cut	uncut		cut
X137	cut	uncut	cut	cut	cut	uncut		cut
X138	cut	uncut	cut	cut	cut	uncut		cut
X139	cut	uncut	cut	cut	cut	uncut		cut
X140	cut	uncut	cut	cut	cut	uncut		cut
X142	cut	uncut	cut	cut	cut	uncut		cut
X143	cut	uncut	cut	cut	cut	uncut		cut
X144	cut	uncut	cut	cut	cut	uncut		cut
X145	cut	uncut	cut	cut	cut	uncut		cut
X146	cut	uncut	cut	cut	cut	uncut		cut
X148	cut	uncut	cut	cut	cut	uncut		cut
X149	cut	uncut	cut	cut	cut	uncut		cut
X151	cut	uncut	cut	cut	cut	uncut		cut
X152	cut	uncut	cut	cut	cut	uncut		cut
X153	cut	uncut	cut	cut	cut	uncut		cut
X154	cut	uncut	cut	cut	cut	uncut		cut
X155	cut	uncut	cut	cut	cut	uncut		cut
X158	cut	uncut	cut	cut	cut	uncut		cut
X160	cut	uncut	cut	cut	cut	uncut		cut
X164	cut	uncut	cut	cut	cut	uncut		cut
X166	cut	uncut	cut	cut	cut	uncut		cut
X167	cut	uncut	cut	cut	cut	uncut		cut
X168	cut	uncut	cut	cut	cut	uncut		cut

Table A.1 (continued)

X169	cut	uncut	cut	cut	cut	uncut	cut
X172	cut	uncut	cut	cut	cut	uncut	cut
X173	cut	uncut	cut	cut	cut	uncut	cut
X174	cut	uncut	cut	cut	cut	uncut	cut
X201	cut	uncut	cut	cut	cut	uncut	cut
X202	cut	uncut	cut	cut	cut	uncut	cut
X203	cut	uncut	cut	cut	cut	uncut	cut
X204	cut	uncut	cut	cut	cut	uncut	cut
X205	cut	uncut	cut	cut	cut	uncut	cut
X206	cut	uncut	cut	cut	cut	uncut	cut
X207	cut	uncut	cut	cut	cut	uncut	cut
X208	cut	uncut	cut	cut	cut	uncut	cut
X209	cut	uncut	cut	cut	cut	uncut	cut
X210	cut	uncut	cut	cut	cut	uncut	cut
X211	cut	uncut	cut	cut	cut	uncut	cut
X212	cut	uncut	cut	cut	cut	uncut	cut
X213	cut	uncut	cut	cut	cut	uncut	cut
X214	cut	uncut	cut	cut	cut	uncut	cut
X215	cut	uncut	cut	cut	cut	uncut	cut
X216	cut	uncut	cut	cut	cut	uncut	cut
X217	cut	uncut	cut	cut	cut	uncut	cut
X218	cut	uncut	cut	cut	cut	uncut	cut
X219	cut	uncut	cut	cut		uncut	cut
X220	cut	uncut	cut	cut	cut		cut
X221	cut	uncut	cut	cut	cut	uncut	cut
X222	cut	uncut	cut	cut	cut	uncut	cut
X223	cut	uncut	cut	cut	cut	uncut	cut
X224	cut	uncut	cut	cut	cut	uncut	cut
X225	cut	uncut	cut	cut	cut	uncut	cut

Table A.2 Resequencing Primers

Gene	Location	Primer Sequence	Direction
dome	322868	GCGCGCATATACGTCCATA	reverse
	323949	CCATTCCACAATCTCGGTTC	forward
	324467	GTCCAGACTCGTCCGTCAG	forward
	328272	GACGCCTGTTGTCTGCTGTA	reverse
	328437	ACTGGCGTGCATGTGTGTA	reverse
	329307	TGAAGCGCTTGTAGTTGTCG	reverse
	329494	CTGCCTGGACTACGACTTCC	forward
	330487	GTCGACAGGTAGCCCCAGT	forward
	330520	GTTTGGCACCTATCGCATTT	forward
Dredd	103613	TGACGAAGTGGTTGTGAGGT	reverse
	103626	ACCCCAATAAGAAACCTTACAAT	reverse
	104818	TCTCTTGCTTGACTGCCATC	reverse
	105060	CAGGAGATCCACTTCGCTTC	forward
	105694	ATAGCCGTGGCCTGAAGAG	reverse
	106354	TCGAATTTTTCGCCAGTTTT	forward
	106399	AAAAGAAGGAAACACCCCAAT	forward
Dsor1	198092	AATGAGTGGGGTGGGAGAGT	forward
	199438	TCAAATCCCATCCATTGCTT	forward
	199836	GTGCGGGAAATTACATTCGT	reverse
	201052	AAAGATAATCCTCCAATGCAAA	reverse
hep	152490	CACAGCCAAGCATAACAGGA	reverse
	152909	TTTGAATTGTCGCTTGTTGC	reverse
	154130	AATCTGCTGGAGCTGAGTGG	forward
	154513	GTGGCAAACTGTCGCTTC	forward
	154506	AACATAGGTGGCAAACTGTCG	reverse
	154642	CTGCACCATCACCATGAAAC	reverse
	155193	CCGCTCCAAAGTGACCAG	reverse
	155406	AGCACGAATCCGTTTCACAG	forward
	156340	TGGCTGATTGCATGAAAAAC	reverse
	156532	ACTGGAAGTGCGATCGGTTC	forward
	156340	TGGCTGATTGCATGAAAAAC	reverse
	156865	TGAGTGAGTTTTGCGTGTGA	reverse
	156998	GGAAAGCCATCATGAGCAAA	forward
	159226	GGGCTCTGTACAAGCGACA	reverse
	159965	ATGTTCGAGGGCTTCACATC	reverse
	160154	GCTCAAGCTGTCCAAGAAGC	forward
	161016	CAATGTCCGATGAACGAATG	reverse
	161527	TTGCTCGCTCATTATGTACCA	reverse
	161631	TAGTTTCCGCGAATTTCAGG	forward
	161638	TGGTTAATAGTTTCCGCGAAT	forward
	162286	CGATTTTTCTCAGCCCACTT	reverse
	163439	TCAGCGACAAAACAAACAGG	forward
	163959	GTTTCTGAGGTGCCGATGTT	forward

Table A.2 (continued)

hop	274275	CCATCCCTTTCGTTTTCGTA	reverse
	275322	GGCGACTGGTGTCCATCT	reverse
	275449	TACGACCTGATGCAGCTGTG	forward
	275469	CCAGTTGTCCCGATTTCATT	forward
	275804	ACGCTTGCTTTTCGCATAGT	reverse
	276983	CGCAACGAGTAAGTTGAGCA	forward
	277393	ATGACCCAACCGAGAAGATG	reverse
	277510	GATCCGAATTCGTACGTGCT	forward
	277896	ATCGAATCTGCGCAAAAGAC	reverse
	279132	CAGTGCTTGAAATGCTTGCT	forward
	279489	CCTTCTCCGTCTGAAACTGC	reverse
	279616	TCTGCAGTGGATCCTTGTTG	forward
	280013	AGTGCAACGGAATTGGTGTT	reverse
	281669	GAACTAGAACCTCGCGTTGC	forward
lz	235452	AAACGATTGGATTCGACTCAG	forward
	236290	TTTGCACTTCACTCGGCTAA	reverse
	237836	TGTCCTTCAAAATCAAAGTGAA	forward
	240167	CGGGTGCACAAAAAGAAAAT	reverse
	244400	GTTAATCGAACTGCGCGATG	forward
	245283	GCGTTTTGGGTTACCGATT	forward
	245450	TTGGAAAGTGGGGATTAGGG	reverse
	246318	AGGGGAAGCCATCGATGTAG	reverse
	253049	GCACCTGCAACACCAGATG	forward
	254073	CAACTTGCAGATATTTTGGGATT	reverse
mxc	194219	CTTTTCGCCTTGCCTTTCTT	forward
	194244	TTTCGCTTGCAAGACTTTAGG	forward
	195025	CAAATGCCTCTTCCTTTTGC	reverse
	194897	GCGACATCAGCGGAGAAA	forward
	197005	ATCTAGCACAATTCTTTGATCGT	reverse
Ntf2	113279	CAAATGGCAATGAATATTTAATTTTAG	forward
	113327	TGCTTTTCCGAATGTGAAGA	forward
	114294	TTCTCGAAGTTTCGGGTGAC	forward
	114483	TGAATTGATTGAAACTAATGAAACA	reverse
	115273	CCCGTTACTAGTGCAGTTAAAGA	forward
	115410	TCCTGAGATCTCGACGTTCAT	reverse
	115577	TCAAAACAAGAGAGAATGCTATGG	reverse
	115955	CGGCAGTTTCTTTGTGCAG	forward
	117037	GCACTGCAAAGGAATGAAATC	reverse
	117613	AAACTTATGTAGGCGATGATCC	reverse
os	152853	CCTCAAAATGGGAAGTACGAA	reverse
	152869	ACGAAGTTCTTTTCCATCATAAA	reverse
	154156	GCGTCTCGAGATGAACAAGC	forward
	154028	TGCTTCACAAAGCGCATATC	reverse
	155238	CGCAGAAGAGAAAGTGGCTA	forward
	155085	CGCTATCGATAACCGTTAGACC	reverse
	155849	CCCGCCCTCAATATACACAC	forward

Table A.2 (continued)

PGRP-LE	95395	AAACTGCCAGTAGCTGGAAAA	reverse
	95994	AGCTGTGTGTACTGCGTGGT	reverse
	97097	TAAGGTGGACACGACACGAA	forward
	97607	TTTTCGGATCTGGACAAAGC	forward
PGRP-SA	191815	TTTTTCCTCGCCCTCTTTTT	forward
	193314	CGACACATTTTTGTAAATTATGACAG	reverse
phl	273661	GCCTATGCACGCCATCTATT	forward
	273273	TGGAAAGGATACAAGCCAGAA	forward
	275609	CTCGAAGCCACCATCTTCAT	forward
	275729	TGTGACCGATCGAATGTTGT	reverse
	276530	TCCGTGAAGATAGGCGACTT	forward
	276690	AAGGCATAAACGTCCGACTG	reverse
	277575	GAACCTTGTGTGTAAACATAAACCA	reverse
	277615	TGAATATACGGCTGGGTGGT	reverse
Pvf1	74559	CGGATTGGATGTGAGTGTTG	forward
	75186	GTGTTCTGGTTTTGCTCGAA	reverse
	76087	CTACTGCTCCCCCGTCTACA	forward
	76833	ACCACAGGGAGACGGATAGA	reverse
	79380	CGTACCCTCGGAGTGATGAT	forward
	79800	CTTCGCGGCTTTGAAGAAT	forward
	80882	TTTCGCTGCGTGAGAATATG	reverse
	80713	CGTTTTGTTTGCAGCTTGAT	forward
	81346	CGGCGAGAAATACTCGTATGTAA	forward
	81484	AACCGATTCCCCACTTGTTT	reverse
	81858	CGTTATTATCACGGCGGTTAG	forward
	82943	TGGTACACGTTGCAGTCCTC	reverse
	82768	GGATCCACGCTGATCTCCT	forward
	83430	GGAATTCGCGCAGTATGAAT	reverse
	83287	CCCAACAAGTTCCTCGCTTA	forward
	84226	CACCCATCGATGTCTTGGA	reverse
	84386	CCTTCTCAGTACTCGGGGATT	reverse
	85062	CGTGGAACTCGACTACAGCA	reverse
Rps6	100048	GCTGGCTAGCTCATAACCAAA	reverse
	100808	GTGTCCCTTCTTCAGGAGCA	reverse
	102386	GGCTTTTCGATAGGCTTGTG	forward
Ser7	236140	TCAATAGCACCAATGCAAAGA	forward
	236901	GAATTCCACCGAGTTGCAGT	reverse
	236787	CAGCAATTTGGCAGTGAGTT	forward
	237893	TTATCACGTCCGGTGAACTG	forward
	238577	AGGGATTTGGAATGCTTTGA	reverse
Tak1	196554	TTAAGGTCGTCGGCAATAAA	reverse
	196578	GGTCGTCGGCAATAAATAGAA	reverse
	196911	AAACAAAGTGCTATGGTTAATCG	reverse
	197877	TTTGCTGCAATGACAATTCC	reverse
	197098	AGAAGGCTGGGTGGTCATC	forward
	198605	TGTGTTGGAGTGTGGAGCAT	forward
	201271	AAAAGATTCCTTGTGCATTCG	reverse
	201992	CAGCACGAACGGTGAGTTT	forward
	201849	GCCATTGTGTCTCTGAACGA	reverse
	202939	TCTGCCCGTCTTTAAAACCT	forward
	202794	GACAATCTGCGAGCAGTTCA	reverse
	203374	CCGTATACTTGGAGCCTTCG	reverse

Table A.2 (continued)

	203437	TCATCATGGTCGACTTGTCC	reverse
	203499	ACAAGGGACGCAATCTGAAG	forward
	204115	TTCTGCGGGGAAAAAGATTA	reverse
	204256	ATCACAATGCGTTCGTGTGT	forward
	205112	TGCGACTGACAACTTTTACCA	reverse
	205249	ATCCACTAGGGTTGGCATCA	forward
	205727	TTATCGTTATCGGCGAGTCC	forward
	205767	TCCATTGATGTCGCTGGAAT	forward
Traf2	38316	GTTCCAAATTGCGCCATAAT	reverse
	39038	CAAAAGTCAATGCAGATCACG	forward
	39572	GATCCACCTTGGTGCAAACT	reverse
	39688	GATCGGCTGCTCATCAAGA	forward
	40081	CAATCGTGCCATTGCTGTAG	reverse
	40916	ACTGCTACTGGCCGCAAT	reverse
	41011	GCGCACAGTTTGCAGCAT	forward
	41679	GATTCGATTTCGCTTTGAGG	forward
Traf3	81982	ACTTCAATCCCGATCCACAT	forward
	82520	TTTGGCTGAGTTTAGTGTGCAT	forward
	83247	TGGATGTCCAGTACTGCTGTG	forward
	84911	CATACAAGAACACGCCAACG	reverse
Tsf1	234903	TCACTGCAATTTTTCCAGCTT	forward
	235473	GGGAAAGAAGCAGCACATCT	forward
	236278	AGCCAGTGTGGCAGGACTT	reverse
	236581	GCCTTCTCCAAGGTGCAGTA	forward
	236694	GCCATCCTCGCACAGATATT	reverse
	237205	AGGGCTCAGGATGATGTCC	forward
	238054	GGAAACTTTGTAGCATTGTATTTGG	reverse
	238059	GGCTGGGAAACTTTGTAGCA	reverse
upd2	88810	TGGTCCTCTATTTGGCTTGG	reverse
	89315	AACTCGATCTCGCAGAGCATA	reverse
	90057	GCTTTCCTCCATTGCCCTTA	reverse
	90303	CCACAACCTGCGACTCTTCT	forward
	91399	CGGAAGTCGTGAATCGAAAT	forward
upd3	122615	GAGAAAGTTCTTCCCCTCGAA	forward
	123203	CCGATCACCATCCGTAAGTT	forward
	123349	GACGTCTCCGTTTTGTCGTT	reverse
	124190	CCAAATATTGGTCTCAATCGAA	forward
	124343	TACGCTGAAGAAAGCATGGA	reverse
	124817	TGGAGTGGAGTGTTGTGGAG	reverse
	126451	TGCGATTATATTTATATGTGTGCGTA	forward
	126465	TATGTGTGCGTATGGGTTTG	forward
	127477	CAAGAAACGCCAAAGGGTAA	forward
	127624	GCCCGTTTGGTTCTGTAGAT	reverse
	128460	TCCAGCGATCACGTTTTATG	forward
	128576	CGAATTGAGATTCGGATTGA	reverse

Figure A.1 SNPlex Oligos

>dome5UTR_01

GGATCCAAGATGGCGTTGGCGCCACCCCGGCAGCCGGNAAGCANCATGAGCAGCANGAC
GAGCTGCTCCTGGGCCACCATCTTCCTGGTCNTGGTTCTGCTCCCAATCCTA[A/T]TCCTAAT
CCTAATCCCCTCCTTCCCTTGCTGTTAAGGTTTACTCTGGCTCNCTCTCTCTCTCTCTCNC
ACTCTGTTCTCTCTGTCTTTAGCTCTTTTCCTCTCTTTC

>domeExon1_01

GTAGTCCAGGCAGTTGAAGTCGCGCACCAGGAGCGGCCTGGTGCCCACGTAGACCTTTGA
CTTGTTGATCACATACTCGTCGCACATGCAGTGGTAGTCGTTCTCTTGCTC[C/T]ACGGCATT
CGTGTCGCTAAACAGGATGGTGGTGTTGTTCAGGATTCGGATGTGCTTGGAATCGCGGTAG
ACGCGGCCGCCGCCTGTAAAGTACAGCTTCTCCACCGG

>domeExon1_02

GGTTTGCCAATGCCCCAATAATTAGCTGCCAAGCAGCTTACGGCGCATATTCATATTCACA
TATAGGTACTAACCACCATCCCAGTTCCCCTCCTACTCACACTGCTGTTCCGTTGATGCGGT
ACTCGCTGATGGAGTACCGAAAGTCCGG[T/C]CCGTTCAGCTCGTGCTCCTCCAGGNGCTCC
CAGTAGAAGCGCATAGCCTTCTCCGACGAGTAGACATAGAAGCTGCCGTAGGTGACGCGN
GGAGGGCGGCGCGGCGGCGCCGGAGCGGTGGCGAACGCATAAA

>domeExon3_01

GAACGGGCACGTAGCCGCTGAGCGGAATCTGGGACAGCGGCTGGGACATCGGAGCGGTCA
TTGCAGTCGAAGGAACTGAGAACGTGTGATCATTATCCGAGGGACCGT[T/G]GCCGCCCCA
GCTCTTCATCTGTGTNGGCTTGATGTAACCATTGGTGCAACTCTGCTCCCGATCCATTTCGC
TCTCGCGCTGCTGTTGCTGCTGCTCCCGCTCCCTCTCCTGCTCACGATCCCTCTCCCTC

>domeExon3_02

TAAAATGATGCACACCCGTGCATAAATGATGAGCGCTCCGGTATCTTAGAGGACGTGCCG
ATTGTGGGCCATGGCGTTCAGATCCTGCGGCGTTACGTAGCCATTGATTTGCGGCTGGAGG
[C/G]GAGAGTGCTGGTTGCCGCCTCCCGGCGCCCCACCAGCGGCGTGCGTCGGCGATCCCA
CGGTGGCGGCCAGAGGTGGCTTTATCAGCCCAGTGAGCTGCAACTGCTCCATGGTGGTATA
GCCGATGTCAGCCATT

>dome3UTR_01

TTAGCTTTATATTGAAATACAGCGATGTGCGTATTGTGTAAATAGGGTATTAATAGTATAA
GCTCGATCCGCTCGACGCGTGGATGGTTAAGTGGGTAAGTAAGTGTACAGTACGAATATAT
[A/G]TAAAAAAATACTATTGCGTTACAAATACTTTCTGCCTCACACGCATCGCAAAGAATA
CAAAATAAATTACAAACGTAAAACTACATTCGTCCGAGGAGTTTTCACTAATTACTATAAA
CAGCTGCCCGACAGTGGTTTCTAGTTTCCA

>DreddExon3_01

AATGCGCTCTATGATGCCCATGTGATCCACGTNGTCGTACGCNNCCACATTGTATCCCATC
GAGGAAAACACCTCGATTAGTCGTTCTTTATCCACATCCGTGCCATCCCGTCTACGCAGTG
G[G/T]TCGGGCGACAGAAATTTCTGTGGAGAAGGGCAATGTTGTACTTCTTTTTGTTACTCA
ATATTGCAGAGATGAAACCATATTATCCCTGCTAACATTCCGGTGAAACTTCTGCTGGTTG
ATAATCAAGGC

>DreddExon1_01 (DreddExon2_01)

GTTACTCAATATTGCAGAGATGAAACCATATTATCCCTGCTAACATTCCGGTGAAACTTCT
GCTGGTTGATAATCAAGGCGATTCCTGCGTTCTCCCGGGTCAA[C/T]TTAAGAGCATCTATT
TGTGTGGAACAGTACGACTGCTGGTTATCCGATTCAATCTCTTGCTTGACTGCCATCGCAG
CAGTGCCAGCCGCATCTGGCTCCGGAGCGT

>DreddIntron1_01

CACTTTGCGGGAAGTCTGATCGTGTCATGGCCAAAAGTTTCTGCAGAATGTAGGTGGCATC
CGAGTGGTCGTCGCCATAAAGCAGAAAGCAGAGGCCCACCTA[G/A]AAGAAAACGAAGGT
TCTACTGTTTAATTAACACTTTGGNCTAGCGGATGTGTAAAACATTTGAGTCATATTCGAA
AAGTCCTCCCAGGGAGGTGCCATACCTTTT

>Dredd5inter_01

CCGATCGATATCGATAAATGGATATGTTTTCTGCANNAGCGAAAGATCGGGGTGCTACCTA
TTAATTATACTTAAGAAACAGATGATGTAGTTTTTTGATAT[A/C]GAAGTTTAACAACCGGT
TCAACAATCGGTTAATATTTAAAAATGTCGCAGCCAGCGAGCCTTTAGCTTTAAAAGTAAG
CTTGCGTTCACGGTTTTTGGGGAAATACCAAATATACTAGGTAAAAACATCGATTAACTAA
TAT

>Dsor15inter_01

TTAGTAATATCTAATCCTAATATCCTTTCCATTTAAGTTGTGACCATTGGCCGCTTTAGTCG
TAGGAGTGTATTAACGGTATTTGGTAATCGCGACATCTGGCCA[T/C]ACCGCACCGTCTGCG
AAAAAAACCCAAAAGTCGCCGATTTTGGCTGATTATTGCAAAAATAATATTAACTATCCGT
TTGCTGTTCGTGTGCTCGGTTCGGGTGTG

>Dsor1Exon2_01

GTCCGACGAGGATCTGGAGAAGCTGGGCGAGCTCGGATCGGGCAATGGCGGCGTGGTGAT
GAAGGTCCGGCACACGCACACACACCTGATCATGGCCAGGAAACTGATCCATCTGGAGGT
GAAGCCGGC[G/A]ATCAAGAAACAGATCCTGCGCGAACTGAAAGTCCTGCACGAATGTAA
TTTCCCGCACATTGTCGGTTTCTACGGCGCCTTCTACAGCGACGGCGAGATCAGTATCTGC
ATGGAGTATATGGACGGTGGA

>hepExon1_01

CAGGGGCGNGACGCGGGCAGATGGAACGGAACCGGAAACGACAGGGCCGCGCGCACCGG
AGAGGACGATCTGGTCGTGGGACTCATTCTGCGCCTGGAGCTT[T/C]GCCTCCAAGGATTGC
AGCCGACTGCCGATCGTTTCGAACTCAATGGTGGACATTATATCGAGTCGATTTGGTTCGA
TTTTTTATAGTTAGTTGCTAGTATCGTTGGATTTTGGTATT

>hepIntron1_01

AATTTCTATCAATTTAAGCCTNATAAATAATCGAATAACGACNACTTTTGTCTTTAATAAA
CCAATTTTGTTCNGATAAATTAACATATAGACCGTTATATTTTGTTA[G/A]CTATTTATTCAT
TTAACTTTTGCTCGCTCATTATGTACCACGAACTTGAAANNCTATNCAATAGCTGATTNCA
CTGTTAAAATTTGGGTGTGATTTTAAGTACCCCTGAAA

>hepExon2_01

ACCGGCTTCTTGGACAGCTTGAGCAGCTTGTCAAAGCACATGGACATCAGCTCCATGCAGA
TCCACACATCCGGATCGCGAACGAAGCAGCCAAGGCACTTGAC[A/G]ATGTACTTGCAGTC
GTGCGATTTGAGCACAACATCCAGATCCATCAGGATGCGTTTGTTCTCCTCCGCGTTGCCA
GTGCGTCGCATCTGCTTCACGGCGATGATCG

>hepExon3_01

TAGTCTATTTTTCCCTATCATAGCCCTCNCGCTCTTTATACNCACCGCCATATAAGCTGCAC
AGCCGGCGGATCGAGTGTTNGCCTTGGAGTCCACCAGGCG[T/A]CCGCTGATCCCGAAATC
ACAGAGCTTGATGTTCCCGCGCTCATCGATCAGAATGTTCGAGGGCTTCACATCTCGATGG
ATGACTCCGTGCTTGTCCTTCA

>hepExon4_01

GACTAAAGTTGTAACCCTCGCCGTATGGCAAACACGGCGGCTCCGAGTCAAGCACCTTGGT
GAGCACCTCGAAGTCCGTGTTGCATCCCTCGTACGGGGATCGCGC[T/G]GTGGCCAGCTCC
ACCAGCGTTATGCCCAGTGACCACACATCTGCGCGAATGTCGTACTTTGGTTTCTTGGGGT
CGATGCGCTCCGGCTNTAAAAT

>hepExon5_01

AGCGGACGGGTCGTGGGAATAATGCTGAAGGATCTGGCACAGATCACCGGATGGGCTCAC
CTCTGGAGCGTTGGATCGCCATTGGNGCGCAGTCGGTTATCCTT[G/A]ATGCTCTGAAACCA
ATTGGGCACATCTACTTTGGCTGATTCATAAATCCGGATGAAGGGCTGGGCCAAAAGCTCC
GGATACTTGGGTCGATCCTGATGGTTCTTT

>hepExon5_02

ACAGGCACAACANATCATCAGCCGGGAATAGCGGACGGGTCGTGGGAATAATGCTGAAG
GATCTGGCACAGATCACCGGATGGGCTCACCTCTGGAGCGTTGGATCGCCATTGG[C/A]GC
GCAGTCGGTTATCCTTNATGCTCTGAAACCAATTGGGCACATCTACTTTGGCTGATTCATA
AATCCGGATGAAGGGCTGGGCCAAAAGCTCCGGATACTTGGGTCGAT

>hepExon7_01

TTGCTTGCGCGGCAACGAGGGGGAGGGTGAGGANAGGCCCAAATTCAAGGTCGTCGTGGT
CGCATTACTACTACTATCACAGCTGTCGNAGGCTTGGGAATTTTCCG[T/C]TGGCAGGCGCC
AAGTGGGCGTGGTCGTTGGCGTTGTCNTGAGCACTGTAGCTGCAATATTTGGCGGTGTTGT
TGCTGTTGCCGCCGCNGCTGCTGCTGTTGCTG

>hepExon7_02

GTGAGTTGCAGGGTCGGCGATTCTGTGGGGAATTGCTTGCGCGGCAACGAGGGGGAGGGT
GAGGANAGGCCCAAATTCAAGGTCGTCGTGGTCGCATTACTACTACTATCACAGCTGTCG[
T/C]AGGCTTGGGAATTTTCCGNTGGCAGGCGCCAAGTGGGCGTGGTCGTTGGCGTTGTCNT
GAGCACTGTAGCTGCAATATTTGGCGGTGTTGTTGCTGTTGCCGCCG

>hepExon8_01

AGAGGTGGAGTGCGTTGATGGCGGCGGCGGCACATAGTTGCTGTGAAACGGATTCGTGCT
GGTCACTCCAACGGGCGGCTGCTGGCGCTGCTGCTGATACCGCTC[A/G]GCCGCCTCCTTCT
CGCGCAGCTGATTCTGCTGGTGGTAGAAGCGCTGCAGCACTGCATATTGAAATTNATATGT
TAGTTNCANACATANCCCAGCAAATGACTAGATCATTATACCT

>hepExon8_02

CCATAATCCGACTGCTGTTCCGGTGATCTAGNCTGGAGCGTGGGACTNGTTCCCGTTGCCT
CGGCGGCGATGGGCAGNGGCTGATACTGCAGCTGCTCGCCAGTGCCGCCCG[C/A]TCCAAA
GTGACCAGCTGCCCCCAATCCCGATGCCGGAGCCGAACCGATTCCCAATCCCGCCATGTTT
CCCGTTCCGGAACCAGACGATGGNGATATGCTGGCGGGATTGATGG

>hepExon8_03

CGCTCCGCTGCTGGAGCTCTGNCCCAAAAGTTGTCGGCGTGCGTACAGCTTGCTCAGCTTA
CTGGCCACCATGTTGCCGTTTCCTCCATAATCCGACTGCTGTTCCGGTGATCTAG[A/G]CTG
GAGCGTGGGACTNGTTCCCGTTGCCTCGGCGGCGATGGGCAGNGGCTGATACTGCAGCTG
CTCGCCAGTGCCGCCCGNTCCAAAGTGACCAGCTGCCCCCAAT

>hepExon8_04

AAGAATGTTATGATTTTATTTCTGAGCTTGTGGACTAGTTAGCAATTCAAGTGATGGCTCA
CCTGCATCGTGTTGCTCCCGACTGCAACCGTCCAGGCT[G/A]CTATTGCTCGCTCCGCTGCT
GGAGCTCTGNCCCAAAAGTTGTCGGCGTGCGTACAGCTTGCTCAGCTTACTGGCCACCATG
TTGCCGTTTCCTCCATAATCCGACTGCTGTTCCGG

>hepExon9_01
AAGCTACTGGTGCCACCGCCGGGCAGCGGTTCCGTCGTGGGCGAGGTGCTCTGGCTGGTAT
TNNTACTGCTGCTGGCCGCCGACAAGGTGCGATAGTAGTAAGACTTGCCACCGCTGCCGTT
ATT[C/T]ACAACGCTCCTGTACGCCGGCGGCTGCGGATGCACTGGCTCCTGGTCTCGATCCA
TGCTGGACGCCACCGGCGACGTGGCTGTGGAATTCAATTGGGTTTTAACATAGGTGGCAAA
CTGTCGCTTCATGGCACCGGCAAGAGTGTTGA

>hopExon1_01
GCCAAATGGGATACCTGATGCACATAGTGCCACTTCAAACGCTCCACACTCAGCGAATTCG
CTTTTAGACTNCGGAAAACCTCACGTATCTTGGAGCCCACAAAGAAGCTGTGGGCGCGCC
ACA[A/G]GCTCGGTGGCAAATACAGTTTGTAAAGCTTCTCGATGGAGCGCATCGCCTGCTG
ATCCTCCGACTGTTCCTGATCGTCGATTAGCATATCCATCACGGCCAGTCCCGTTGATTTGT
CC

>hopExon1_02
AAAAGACCAAGCAAGTTCGGCTTAAAAGATTAAAATTTAGTTGGAAGCGAACTCTTTGGG
ATNTAAAGNTCACCGATTACAAGTGAACCACTCACCTGATGTAGAT[A/G]CCCAGGTAGGA
GATGAAGGACTTCATCTCCTTCTCCGTCTGAAACTGCATCTCGTAGCCCTTGGGCAGACCA
ACGATTTCCAGCCGCACAGAGG

>hopExon4_01
TGATGTGCGCCACTTCGTGCAGTGAGTTTANTACATGCTCGCCGTTGTTGTAGCTTAGCTTC
CATTGCGAGTCCTTGCGCACGATCCTAAATGTCTCCGTCTTGCA[G/A]CGTTCCTGATCGGT
TTTCTTGGCCATGCTGCAAAGGTGAGCGGGATTTTCGCATGAGATAAAACCGATCACATCT
AATTNGGGCATCATCAATCACACACACA

>hopExon4_02
GCAGTAGCAGCAGCGGCGGCTTGTCATATTTGGAGGCGGGNATGCNATAGCGATCGGGAC
TATCCGCTTGAATGATGTGCGCCACTTCGTGCAGTGAGTTTA[G/C]TACATGCTCGCCGTTG
TTGTAGCTTAGCTTCCATTGCGAGTCCTTGCGCACGATCCTAAATGTCTCCGTCTTGCANCG
TTCCTGATCGGTTTTCTTGGCCATGCTGCAAAGGTGAGCGGGA

>hopExon7_01
AGACAGTGCCATAATGACCGCGCCCGATCATGTTCTCCATGTTGTAGATCACTCGACATTC
GCTGGTCAGCGGTATGACCATCAACATATTCGAACGCGGAAAGGGTAG[A/G]AAGGGTAT
GTCGCTTCTATCGATCACGGTCTCGTCTCCAGTTCCATNTGGAAAAAAAAATAGAAGATGA
AATGTATGCCAATCTCNTAACGCTGCTTTCGNCTGCTCA

>hopExon7_02
AACCACTCACATTTGCAATATCTAAAGCGAAGCTGACGAGCCGTGGATTATTGAGATTGGG
TGCCGTGAAACGCAGGTAAATGTCGAAGGAACCAGATTGCAGATA[T/C]TCCATAATGATG
CAGTGCGACTTCTCAGCCCAATACTTGAACTTGACTATGTTCGGATGACTGAGGGTGCGCA
TGATGCCGATCTCGCGATGAAAATCGGTCGACACCTG

>hopExon8_01
CGGAANTGGCGAANTGCGCCAGGCCGANGTCTGAGATTTTGACGCAGTCACCATCGCCGT
TGTGGTCCACCAAAATGTTGCGGGCGGCCAGGTCGCGGTGGAT[G/T]AGTCCCATGTCGGA
CAAGTATTTCATGCCCTGCGACGAGAANCATTTGGATTAAGTACTGNATCAAAAGGAAAT
ATNNCTTCTAAAGTCGTAAACATGAGTTTTCTATACCAAAG

>lzIntron1_02 (lzIntron2_01)
GGGCTTCCACATCGTGGCAATCACGATCATCGGACGCAGCCTTATCGGATTCCATGGGTAT
GATGATCGTGATGATGATACCGATGGCGATGATGATGGTGAT[G/A]CCAGTGAAGATGGGA

CTGGCATCGCCGTCGGAATATGCGTAAATTGATATCAATAATTGTTATTGATATCAACGCA
CGAGNCGTTCCCCTTTTTTATTTTGTTA

>lzExon5_01 (lzExon6_01)

CGGTGCTGCCCGAGATGCATGGCCATGGATTCGCCACCGATCCCTACCAGACGGCCGGCTA
TGGCGGTGGCAATACGGGCGGAGGCAGTGCATCCAAATCGGAACTGGACTA[T/C]GGCGG
CAGCTACAATCAGGCGTGGTCGAATGGCTATCAGAACTATCAGTATGGCAGTTGTTTGGCC
ACCGCCCAGTACGGTCCTCAGG

>lzExon5_02 (lzExon6_02)

TCGGAACTGGACTANGGCGGCAGCTACAATCAGGCGTGGTCGAATGGCTATCAGAACTAT
CAGTATGGCAGTTGTTTGGCCACCGCCCAGTACGGTCCTCAGGC[G/T]GCGCCGCCACCAC
AGCCACCGCCTCCGCCGCCGGTGGTGCTGTGCCCGCAGCTGTACTCCACGGTCAATCAGAA
TCAAATCCACCTGCATCTGCACAGCAGCGAGAAGCTG

>lzExon5_03 (lzExon6_03)

ACTCCACGGTCAATCAGAATCAAATCCACCTGCATCTGCACAGCAGCGAGAAGCTGGAGC
AGTATCTGGGCACGGCCACTAGTGCGGACCATCTCACCATTGGATC[C/T]CTGACGGGCAG
CAGTCGTTCCAGCATTGAAATTGGCCAGGATCAGTACCANCAGCAGGTGCANCACGCACA
GCAGCAGCAACAGC

>mxc5inter_01

TATCGATATAGTCCCGTTGTTTGCGCTTGGCTGCCAAGCCTTATTAGGAGCGNGGGAAATA
ATGGTAATAATTAAAGCAAAGAAATACGTCTCTTTCGCTTAAAA[C/A]AAATTATAGCAGT
TTTTATTGCTTTGANGGTCTTGCTTCAATTCGTTTTTATGTGTACGTTACAGTACTTAGAATC
CGATATTCGTTAAACTATCGATATGGGAATGCTGC

>mxcExon2_01

CATAACNCTGATCATGTTATTCTTTTACAGATGTACCATTGGAAATCTTTCTCACTTTCAAT
AAGATAAAGACGCTCACCCAGGATGTGCAGCAGATAGCCAAGGC[A/T]CTAAGCAACTCCC
AGCTTCTGGAGCTGGATGAGACCGAACTAAAGGTCAAGCGACGTACCAAGCTACCAGATC
AGCGCGATGTCAACGACNAAACCTTGTA

>Ntf25inter_01

GTGAAAATACAAGCGATTCCACATTATGAATAAACGGATCTAAAAATCAACTTTAGATAC
CTTACAATGGATTGTAAGTTGCGTTAGNTTAAGTATGACCTG[C/G]AATTAAACAACAAAT
CAAAGAAAAGTCGTCCCGAGCGAGGAAAAAAAATAATATATGAAACACAAATAATTGTTT
TCCAAATGACCAAGTATTACGCCTGACAG

>Ntf2Intron1_01

AAGTTTCGGGTGACACATTACGCCCANGCNCGCATAAATAAATACNCACATACGAAATAT
ATGGCCAACATACAACTGCGGTNAGCACNATAGGGCCCAATATAC[G/A]ATATAAATTCGT
GAAAAAATTGTATTAATATCTTGAAAAATATGGCATAATTGTGTTTCATTAGTTTCAATCA
NTTCAGTTNTTTTGTCACAT

>osExon1_01

AGCTCTGTGGAGAAGCAGATAAAATGGGGAGCGGATTACATATGGATCGGGTCCAAAGAG
CGCTGCCGTTTGAACTCACCTGCCTGGGACGCTGGCGATTGAAGCGATCACCCGCCGAGGG
ATGATAGTTGCCACCGCAGGGATTGG[C/A]CCACTCCAGGGCGGTGGTGTTCTCGTCGAAG
TTGCGATAGTCGATCCAGTTGCTGTTCCGCTTTCTGTGGCGCTTGCGCGGACTAGCGGCGCT
GGCAGGGATCGTGCCCTGATTGGGATG

>PGRPLE3inter_01 (PGRPLEExon3_01)

CAGCAAACTTGAATACTCACAGATAACCACATATTTAACAGGCAGTTGCAAGGGCAGTGG
CTCGTCCATGGGCTTCTGTGCTAGCCACGAAGAGCGCGGAATGATGGCG[G/C]ACAGCTCC
TTGGGTATTTTGAATTCTGTAAGTGAAATCGGTATCCTATGGTTAACGTGTATGCATCCCCT
GGTGAATGATAGCTTACTCTGGCGCCTGA

>phl5UTR_01

CGACGNNNTGTGTCGGGGCTGCCACTTGAATCNAACCCATNTATGTTTCTTCCATCACTAC
CNCCTGNACCTTGTATATGGTTAGTTGATTAATAGCCACGTCA[G/A]AAACTAATTTACCTG
TTGCCGCTCGTACCAGATCCAGATTTGTACTATCCCGAGAAGTTAAAAGCTCTAGGCAAAT
TAACAATTAGCCGCGACACAAACCCCGTTTC

>phl5UTR_02

ATTCGCGGAAAGTAAATAAATTGTTATAGCCAAGGTGAAATAACGAGCGGCCAGCTAGTG
GCGATACTGATACTNTTGCGAACGTTGGGCAGCCACCGAC[G/A]GTGCCGGCTGGTCAGGN
TGTTATCGGGTAATTGGCAGCTCCTTTGGAAAATCCTCAAGTTCAGCTGCTTCTGCACACA
CTGACCTTCATTATACATACATACCGTATATA

>phlExon2_01

TGCATGTGGAGGAGATCTTTGTCAGGCTGCTGGATAAGTTTCCCATTAGGACACACATCAA
GCACCAGATCATACGGAAGACCTTCTTCTCGTTGGTATTCTG[C/T]GAGGGCTGTCGAAGGC
TTCTGTTCACCGGGTTCTACTGTAGCCAGTGTAATTTTCGATTCCATCAGAGGTGTGCCAAT
AGAGTGCCGATGCTGTGCCAGCCCTTTCCCA

>phlExon2_02

CTGTAGCCAGTGTAATTTTCGATTCCATCAGAGGTGTGCCAATAGAGTGCCGATGCTGTGC
CAGCCCTTTCCCATGGATAGCTACTATCAGCTACTGCTGGCCGA[G/A]AATCCGGATAATG
GCGTTGGTTTCCCCGGCAGAGGCACTGCTGTCCGCTTCAATATGAGCAGCCGGAGTCGCAG
TCGTCGTTGCAGCAGCAGTGGCAGCAGCAGCAGCT

>phlExon2_03

CCGGATAATGGCGTTGGTTTCCCCGGCAGAGGCACTGCTGTCCGCTTCAATATGAGCAGCC
GGAGTCGCAGTCGTCGTTGCAGCAGCAGTGGCAGCAGCAGCAGCTC[G/A]AAGCCACCATC
TTCATCCTNCGGCAATCATCGACAGGGTCGTCCGCCGAGGATCAGCCAAGACGATCGTCCA

>phlExon4_01

ACACCGAGTCCCGCCCAGTTGCAGGCGTTTAAGAACGAGGTGGCCATGCTNAANAAGACG
CGCCACTGCAATATCCTCCTCTTCATGGGCTGTGTATCCAAACCATC[T/C]CTAGCGATTGT
GACCCAGTGGTGCGAGGGCAGCAGTCTCTACAAGCACGTCCATGTCAGCGAAACCAAGTT
TAANTTGAACACGCTCATCGATATCGGACGTCAGGTGGCC

>phlExon4_02

TCTCTACAAGCACGTCCATGTCAGCGAAACCAANTTTAANTTGAACACGCTCATCGATATC
GGACGTCAGGTGGCCCAGGGNATGGATTACCTGCATGCCAAGAA[T/C]ATCATTCATAGGT
GGGTTTCCGTATGGTCANCAGTTGTAATCGGTTAANAATNATATTTTCATTTCTCTCTTTTA
GAGACCTCAAGTCAAACAACATCTTTTTGCACGAGGA

>phlExon5_01

GGCCACTGCGAAAACTCGATGGTCGGGTGAAAAGCAAGCCAATCAACCCACGGGCAGTAT
TTTATGGATGGCTCCAGAGGTGATTCGCATGCAGGAGCT[A/G]AACCCCTACTCCTTCCAGT
CGGACGTTTATGCNTTTGGTATCGTGATGTACGAACTGTTGGCGGAGTGCTTGCCCTACGG
TCATATTAGCAACAAGGATCAGATCCTGTTTATGGTGGG

>phl3UTR_01

ACTACTGAAACTAAACTAAACTAAACTAAACTAGCTGATCGCAATTACATTATACACATTA
TACTTATACTACAAGAGATGGTGTTGTTTCTGGAGTCGAGC[A/G]CGATGAAGAACATTTA
ATTCAGGTATTATTTGTTATNCCATTGTCATTAACTACTTCTTGGTTTATGTTTACACACAA
GGTTCTAGGAATTTTTAAGTATGGAACCACCC

>Pvf15UTR_01

ACAAACTTAAACGCGCTCAAATATATATCACAAATTATTAAATACAAAACCCCCCATTTGT
GCACACATATTTTTTGTTGTTTGTTTTTCGTTTCGTGCTAACCGATTTGTGCCATTGAACTTC
GTGTAAAACATAATATTT[A/T]GAATTTTCAAGCCGTGTTTCCATAAAGTAAGTGGTGCTTC
TCAAAATTTCAACATTTCAACACAAATCAGCAGCGTAAACAAACAAGCNAGCGACCCTCG
AC

>Pvf15UTR_02 (Pvf1Intron1_01)

GTATTGGAATTCTTGTGCGGGGAGCAATTGCCTGGCGGGAGGCACTCCCTTCGCGTAATTA
AGGACCGCCCAAGTGGGGATCGCGNTNCCGAGATGATTCCTTT[T/C]GGTTCCTCGCCTTCT
CCTCCTTGAGATTACTTGATATGATTTGATATCTTCGGGTGACGCACCGGGAAGGAATGCC
TCGAAAGTGTGAGAATAAATACAGCTGA

>Pvf1Exon1_01 (Pvf1Exon2_01)

GAAAACTGTTCTTCCATCGCATCCGCATGGCCATGTCCCGAATCCGAATNCTGAAGCCCAT
GCAGCATCCATTGCTCNTCCTGCCAGTGCTCCTTCTCGTCCTGATCGTCAGCGAG[G/T]CAG
CGGCGGGCTCCTTGGTGTCCCCGAACAACAGGCANCCNTCGCAGAGGTTCTTCTACGCCGC
CGCNACGTCCTCCTCCTCCACGCAGACCAAGGTGCGCCTGCT

>Pvf1Exon1_02 (Pvf1Exon2_02)

GCANCCNTCGCAGAGGTTCTTCTACGCCGCCGCNACGTCCTCCTCCTCCACGCAGACCAAG
GTGCGCCTGCTCCGCCANGCGGACGATCCCAGTGCGGCGCCGGGT[G/A]CACTGGAGGGCA
GCTGCTGCCAAGGAGCAGCTGCGGAAGCCGGAACACCATTGACGGTAAGACAAATGCTAC
AATCACNCCACACATCCATCANTATGGTAA

>Pvf1Intron1_01 (Pvf1Intron2_01)

ATCTGAGACCCCAATCCCAGACCCCAAACCCCACCAAAGATCGGGAGCAGCCGGTGGACA
GTGGCGCCGTTGTCTCAGCATATTCTCACGCAGCGAAATGGGAATAGAAACTACACACCA
A[A/G]AGTACAGTTTTAACTAAATTTTGTAACTAGCAAGNNAATTTTATNTTTNTTTTATTA
AAGNAGNAACTTGTAGATACTTTTGNAANCAATTTACTAAGTA

>Pvf1Intron1_02 (Pvf1Intron2_02)

TGCTCCTCCGCCGCTTGGAACGGCTACCGAAATACGGTCAAAATAACTGGGTGACGCACG
ATGGCGTCGATTTGGTGTTNATTCCATCCATTTNTATGCAGTCGCCGTA[C/A]GTTTTTCAAT
GGAATCTCGACGTCGATACGCATGGTAGTAGTGGTAGTAGTGGTAGTAGTGGCCCATGACT
AAATCCAAATGAATTTGCGGNTGAACTGAATCCATGT

>Pvf1Exon4_01 (Pvf1Exon5_01)

ACCTAATATCCCAATTACNATATTCCCAATCNCCAGCAACNGTGAGGAATGCAACGCCGG
CGAGCTGCTCCCCGCAACCGACGATCGTGGAGCTGAAGCC[A/G]CCGGCGGAGGACGAGG
CNAACTACTACTATATGCCCGNGTGCACGCGGATCAGCCGATGCAATGGATGNTGCGGAT
CCACGCTGATCTCCTGNCAGCCAACGGAGGTA

>Pvf1Exon4_03 (Pvf1Exon5_02)

CTGCGGGTGCGCAAGGTGGACCGGCGCGGNCACNAGCGGACGCAGNCCGTTCANNTATCA
TCNCCGTGGAGCAGCATACGCAGTGCCGCTGCGATTG[T/C]CGCACGAAGGCGGAGGACTG

CAACGTGTACCAGTCGTATCGCAAGGATNTNTGCCGCTGCGANTGCCACAACACNNANGC
CCGGGACAAGTGCCTGGAGCAGGCCGAGA

>Pvf1Intron4_01 (Pvf1Intron5_01)

TTTTTGCAGTTAAAGTAGAAAATGAATTAGTGAAAAACATTTAGTAGTTATCAGTTCAGAG
TTCTCAGGTNCCACTGTTTGGTTTTTGGTTTTTGTTTAAACTTATTATTTGTTCAACTGCT[G/
A]GCACACACTCTGTACATATATATAGTTTACGTAATTTANGNCACTTTATAAATGTATGTA
NGNTTTTACTTCNGTTTCTTTTGTGTTTCGTTTTAGATTAAAAA

>Pvf1Exon5_01 (Pvf1Exon6_01)

ATAACCTAAAATTGTTAATTAATTTTCGATTTTGTTCTTTCATTCGCAGCCGGCTGCACCCA
TCGATGTCTTGGATCGCAAACGCTTCATTGTCCAGGCGGTACAGGTGGA[T/C]CCCGATAAC
ACCACCCTGTACAGTGTTTGAGGACAGTTGGTTCGCCNCGGAGNNCGAAGTAGAAACGCC
AGCNTCCTAGCGGGGAATCCCCGAGTACTGAGAAGG

>Rps65inter_01

ACACGAGAATGGAGATAATTGCACTGATTCATCAAACATGCTATATTGTGCGTTCTAAATA
TTTATTTATGCGACTTTTTAAATACATCTTGAAACGAATTATATCG[G/A]TACAATCGATTTC
ANNATAAACTTCAAAAACAATATATATAGATTTTANNGAATTDCTGCATTCAATNTCGNNN
NCNCNNNNNCATCCCTGGTCTCACAAGC

>Rps6Intron1_01

CAATCNCTGGACNTTNTCGCTTGCAATTTGTTTGATCNCCGCGCCAATGCTTGCGCTTTAGC
ATTTTCGCAATATTTCACAAATTGATATAATTTATATCAGTA[T/A]TATCACAAATTTTTAGC
NCACCTTCATATTGTCGGTCTGTTTGCNCGCAACGCAGAAAAAAAAAGAAACGGCGGAGG
TCTGTTGGCTGCGAGCGCAGGGTTGAAAA

>Rps6Exon3_01

TTGGACTCCTTCTTGCGCTGCACCAACAGNTTGGCGTAGTCGGCGGAAGCCTCCTTGGAAG
CGATCTGGCGCTTCTTCTTCAGCGCAATGCGACGGTGCTTGCGCTGCAGCACAACGGG[G/T]
GTGATCAGGCGCTGAATTTTGGGGGCCTTGGAGGTGGCCTTCTTGTTGTCCTTGGCGGGCA
AAGGGCGACGCACAACGAAGCGACGCACATCATCTTCCTT

>Ser75inter_01

CTGGAAGCGCGATTGGAAGGCGGCAACGGGTTTTNCCGCCCAGCCGCAATCAATGAGAAT
CGATCCGATCGTAGCTAATCAACGCAGCGTCGGCGATGGAAAAG[G/C]CAAATCAAAAGC
CAATCTCTGAGATTCAAAAAGCTTCCGAGGGTCGCGGCTTGAGGTTGCCGATCGCCAGTAG
ATCAGTAGTTGATTGCCAGCCACCGGGGACGC

>Ser7Exon3_01

TGAACTGGCTGCAGATGCAGAATCTGGAGCCCGTCTGTTTGCCACCGCAGAGGGGCAGGT
ATGCCAATCAGCTGGCCGGCTCAGCNGCCGATGTTTCCGGCTGGGGC[A/C]AGACCGAGTC
CAGCGGCAGCAGCAAGATCAAGCAAAAGGCNATGCTGCACATCCAGCCGCAGGATCAGTG
CCAGGAGGCNTTTTACAAGGACACCAAGATTACGCTCGCCGATAGTC

>Ser7Exon3_02

CGGCAATCGATATGTCTATCTGGCCGGCGTGGTCTCCATTGGACGGAAACACTGTGGCACA
GCGTTGTTTTCCGGAATTTACACCCGGGTCAGCAGCTACATGGACTGGAT[A/T]GAAAGCA
CCATTCGAGCCAATCGCATTTAAGCGACTCTCGACTTTCGTTTCCGTTTTCGGCTATTTTTG
GGCACCCATAACGCAAAGTCAAGCTCATTGTGACCTT

>Tak15UTR_01

GCGCTATTTTGGTAGAATTACNGGGATTTCGCCTTTTTAGTGGCTACGATTCCGTGGGTTCC
CGATTTCATTGCTGGCTAACTGTGCCTTTGGTTCCCTCTTGGG[G/T]ACAGAGTTTTCGGAAT

CTTTGCAAGATCTAGACCTGGATTCATCATAAAGTTTCTNGCCTTTGTATAGATTGCGACTG
ACAACTTTTACCACGTCATTT

>Tak1Intron1_01

TGGGTTTTTTAACNGAGCTGTTACNAAGCATACACTTTTAAAACGAAAAAATAACCGNTTA
TAGATTAAAACATGCTTTNTGAAAGTGCGATAATCGTCT[G/T]ACTAACTAAGAATGAGTA
CTTGCTCTGCCCTAAATGGTATATGATTACTGGTTGTTTATCGTTTCAAGATTGCGTATACG
TATACACCAATAGGAGCTCTAGACTTT

>Tak1Exon2_01

ACGCGCGACAACTGCTTCACCTCCTTCTCGATGTCCTTCTGCTCGGCGCTGGCGAAGAACT
CCTTGACGGCAACCAGCTTGTCGCGCCAAACGGCCTTGCAGACCAC[T/A]CCGTAGGACCC
ATGGCCGACTTTCTGCGGGGNAAAAGATTATTAGATCTGANTCAACAGTTGTACCCTCNTT
GNGCANNACTCCTCAATACNNANTATACGTTATACANCTTG

>Tak1Exon3_01

GCTTCCTGGACAGAACCTCCCATAGAACAATGGCCCAGCTNAAAATGTCACACTTCTCCGT
ATACTTGGAGCCTTCGAAGACCTCGGGCGCCATCCAAGCGGCACT[G/A]CCGCGATTGTTG
GTCATCATGGTCGACTTGTCCGCCACCGTGCCGAAGTCNCATATCTTCAGATTGNNTCCCTT
GTNRRTCAAGAGCAGGTT

>Tak1Exon4_01 (Tak1Intron4_01)

ACAATATCGAATTTAAATTAGGTTTTAAAGACGGGCAGAGTAAAAGAACGCTTTCCTCAAT
AGATGTAAANAATTNACCAATTGACTCTTTTACAATTGTGTATTGACCCAATAAAGC[T/C]G
TCTCACCTGTTGATTAACAAANGTGTATTCCAGGGCCTTGTCCGCCCCCGTATAGTCCTTGA
CGATCTCGTGCATAACGCCCACTATGTACTGCATCGACGGG

>Traf2Exon1_01

GGATGCACCATGCACAACTTGATGCGACCCTTGAAGGGCCAGTCCAAGTGGTAATCGTTCT
CCGACTGCATCAGATGCACATGCAAGCTGAGCACATGCGGTTT[C/T]CGCGGCTGAATGTT
CAGCCGCGCACAGAANTTGTAGCCATGNGGCGAGGTGTAGCACTCGTGCGAGTACACTTG
ATTGTTGGCATTNGNCCGCAGGCG

>Traf2Exon1_02

TATTCTCCAGTCGCGTNTCCTGCTCGCGNNTNNNNTNTTCCAAAACCACAATTNNCTGGNA
CATCGTTTGCACAATCTGCTCATCCACACCATTGGCATATTGCG[G/A]CGGTGGCGGCAGCT
GCTGCTGGCCATGACCNTTNTCCACGGCGGCGCCACTTGTGCTGGGCTTGTGNGGCTGCCA
GGTGGCAATGGCTGTCTGCTGGAATGCCTGCA

>Traf2Exon1_03

AGATGTCGTGCTCGGCGGACAGACGCTTGTTGTCCATGGGGCAGCATTGATTGTTCTTTTG
CATCCAGGCGGTGAGGCAGCTGCGACAAAATCGATGGCCGCACGA[C/T]GTCAACACCGGC
TCATTTAGCCAATCGATGCAAATGGCGCACTCGTATCGCGAGTCCAGCAGCTCCTCCTCCT
GTTCCCCGGATGTATCCGAACCNGGTGCAT

>Traf25inter_01

TANCGTGTTAATTGTGCTAAATTTAGTTNTTATTTTACANNTAANANNAAATCAATGTAAA
ACCATCACTTGTATAGTTTANTTGGTTGTTAGAAGCTGATA[G/A]CGTGTAAACAACATGGC
CTGCGCAGATGCAGGCCAGTGTGAACACNCTCGTTTNGTTTTGGCGGCGGTGGGCGAGCAT
ATAAAACGGCAGTGTTTTTCGGATTGGCGAGC

>Traf3Exon2_01

TCCCATNTGAAGGAGTGTCCACGCAACCAGCACAATCTGAGCAATCAGCAACGCATGAGT
GTCAGCATGGATCGCCTGGATAGGCAGTCGGATCAGCGTCT[C/G]CTGGTCATTGAACAGG

ACGTGGGCACCATTCGGTCCGTTCTCAACGAGGAGATACGACAGCGTTTGCATCTCATTAC
CGATGTGGGCAACATACGGAAGCAGAACCA

>Traf3Exon3_01

TGAGAACGAGNGNCTCAACNGGGAGAAGTTCTTTCAGATCGAGGAGTTNCTGCAGCAAGA
TCAATGAGGACATCAAAACAAAGCTGGGCAACAGTGACTATGTGACGTC[A/G]AAGCAGG
CCACATTGGACTANGAGGTGAAGAATGTGAAGAACATTGTGTGCGAAACGGAGGAACGTT
GCGATAAATTGGATCGAGCACTTCACCAGACCATGCAG

>Tsf15inter_01

TAGCATACTAATTGATTTGCGAAAATCGCATTTAATGGCATAGCCATATATCAGTGTCGNA
AACTCGTTTGCCGGAGTATATAGGCCTGGATTGCCGCCCACTCA[C/T]GCCTTTCCTCGGGG
AAAAACCGGTCGCCTGGACTGTCTTTCTTTCTTTATTGTTATGATTGCTATTNTTATTATTAT
TACTATTATTATTGGTATTGGTATTGGTATTA

>Tsf1Exon2_02

TGGAGTGCGTGGCTGGACGGGATCGAGTGGACTGCCTGGAGCTGATCGAGCAGCGCAAGG
CCGATGTGCTGGCCACCGAGCCGGAGGACATGTACATCGCCTA[T/C]CATCGCAAGAACGA
GGANTATCGCGTGATCTNTGAGATCCGAACGCAGCAGGACAAGGATGGTAAGTGGTGCAN
ATGCTCGTGGGAGCAAAA

>Tsf1Exon3_01

CCTGCCACACTGGCTTCGGCCGCAACGTCGGCTACAAGATCCCCATCACCAAGCTGAAGA
ACACGCACGTCCTGAAGGTGTCCGCCGATCCGCAGATCTCCGC[T/C]ACGGAGCGCGAACT
GAAGTCGCTGTCCGAGTTCTTCACGCAGTCGTGNCTGGTGGGCACCTACTCCACGCATCCG
GAAACGGATCGCCTGCTGAAGAAGAAGTAC

>Tsf1Exon3_02

AGTCGCGTCTGGAGCGCTTCTTCGCCAATGGACTGCAGGCGCAGAACAAGGACGCCGCCG
CCCATCTGCTCATCCAGCCGAATGCCGTGTACCACAGCAAGGATGC[T/C]GCCATCGATCCC
AAGGTCTATTTGGAGCGTGCCGGCTACAAGGATGTGATCGAGCGTGATGGCAGTGCCATC
AGGAAGATCCGCTTGTGNGCCCAGAACGACGA

>Tsf1Exon3_03

CCATCGATCCCAAGGTCTATTTGGAGCGTGCCGGCTACAAGGATGTGATCGAGCGTGATGG
CAGTGCCATCAGGAAGATCCGCTTGTGNGCCCAGAACGACGA[C/A]GAGTTCGCCAAATGC
CAGGCGCTGCACCAGGCTGCCTACGCCCGCNACGCTCGTCCGGAACTCGAGTGCGTTCAGT
CCACCGATTGTGTGGTGGCTCTGACCAAGAAG

>Tsf1Exon3_04

CGCAACNGGCTANGCGGATGCCCGTAGCAACCAGCTGCAGCCAATNGTNTACGAGCAGAG
GGCTCAGGATGATGTCCTTGTGGCGGTCGCAGCACCCGGCGTTACACGGGAG[G/T]CTCTC
CAGAAGGCCAGCATGTAAGTGGAATCCCTCGATCATCCGTAGCANATTAGCTACATATACT
ATCAATATTCCGCAGCAANTTCAATGNGANTTGCGAACGATCCCGTGCTGCTGCCGCC

>upd2Exon1_01

GATGATGAGGATGACGATGACGACGAGGAAGATCGGCCGGAGTTGTCCTCGTCGTAGGNG
TAGTCCAGTGCCACCTGGTCGCGCAAGTGTCGCGCCTTGGT[G/A]AATGGCATCACGACGC
TCAGGATCATGATCACTAGCAGCACCTGCCGGNTGTGGCAGCTCCAGCTGCTGTGTGCACT
CCTGCTGCTGCGGCTGCTTTGGCTNTGTTGGCTG

>upd2Intron1_01

GAAATCGATGAGTAATGCATATTGGCTGTGTGACAGAATTTCTCGAACGTTTCGGGTTCAA
CTAACTGCCTAAGTANCTGTCCTGCNAAAGATGCTAACAGGACAC[A/G]CAACGCACAAAA

GTATCTATGTATCTTAGTATCTATGAAGNTGCNTTGCGACTATGTAAATATAAACTGTTATC
AAACAATGTAAGAAAAGCTTGGTGTAGGTTGCTTTCCTC

>upd2Exon3_01

TCGCCGCAGGACACGGTGCAGGGTCTTCCACACGTTGTTCAGATACTGGAAGTACAGCTCC
TTGCTGACCTTCAGGTCGCGCTGGTCCGCCTCCATGGATCCGTTGGC[T/G]GGCGTGTGAAA
GTTGAGACGCTCCTCCATCGCCTGCCGGCTAACGCGGCTCAGCTTCGCACCATTGCTGTTC
GGATAGGAGGCGTTTATGGTCAACTCGA

>upd2Exon3_02

TGCCGAGGGCTCAAGACTCATTGGATCCGCCATCGGAGCCGGAACTGCCACCGGATCCGTT
CGAGGAGCCCACGTCGAGCATTTCCGAGCTGCCCGCCCGCAG[C/G]CCGCTCGCACCAGTT
CCGTGAATCGCCGCCAGGTGTCGCCGCTCGGCACTATTGCGTCGCGGTCGCCGCAGGACAC
GGTGCAGGGTCTTCCACACGTTGTTCAGATACTGGAAGTA

>(upd3Exon1_01)upd35inter_01

TCCGCCAATTGGACAGATCAACACGATAAGCCGAAATCAAAAGCAGCAGTTACCCGATCC
GTACCAGTGTGCCTCGTACAATGGTTTAAAAATAGCTCGGCCAAATCAT[G/A]ACACCGAT
CACCATCCGTAAGTTTGGCCGCCANCGGTGNGCGTTGGGCGGCNGACGGCGAGTGGGAGT
GGCAGTGGCAGTGAGCGGTACACGGTACATGGTAC

>(upd3Intron1_01)upd35inter_02

CGGTGGAGCGTAGTCGGCAANCATGAATGAATCGCCGTGTCTGCGAGCGTGAGAGCCCGG
CTCGCGGTTGAACAAATCGTGTCGTCATTGACGGCTGTCG[G/T]CGGCCGACGGAGCGTAT
GAGTAATATTCTTGTGAAATTGAACGGAAATAANTGAAATTCATGCCAACCCCCATAAATT
GCCGACTATTTGCGGGGGNTACTTCGGAACG

>upd3Exon2_04

CAGAGCGGCCAACTTCCGGCTGACNTTCCAGCAGAAANTNAATGCCAGCAGTACGCATCT
GGACTGGGAGAACACCTGCAATCTGAAGCCCACGGGTCTGAACGAAACGCA[C/T]AGCAA
GGCGAAACGCTGCAAGAAACGCCAAAGGGTAAGTTTCAANCGAAANTTGGCTCTTANCTG
ANNNATNNDTNGGCACAACTAATTCACACACTATTATCTATA

>upd3Exon3_01

TGCAGAANCTACAGAACCAAACGGGCCGCGAGCTGCGGGGCATCCAGGCCGAGGACAAG
GCCAGGATCACCACCAATGCGGACAAGCTGGCCACAGTGAGCACCAA[G/A]ACTCTGGAC
ATTGTCGAGAAGAACAAGTGGCGATTCTATAAGGGAAACTACAGATTCCTGCCCCGTNTG
AATCTCACTAGCAAACAGGTGAGTGT

Table A.3 Genotypes at Each SNP

Line \ SNP	dome3UTR_(	dome5UTR_(	domeExon1_	domeExon1_	domeExon3_	domeExon3_	Dredd5inter_	DreddExon1_	DreddExon3_
1	NA	A	C	NA	NA	NA	A	C	NA
2	NA	A	NA	NA	T	NA	NA	NA	NA
3	NA	A	T	NA	NA	NA	C	C	G
4	A	A	C	NA	T	NA	A	C	G
6	NA	A	C	NA	T	NA	NA	NA	G
7	A	A	NA	NA	T	C	C	C	G
8	NA	A	C	NA	NA	NA	C	C	G
9	NA	A	C	T	T	NA	A	C	G
10	NA	A	C	NA	NA	NA	C	C	G
11	NA	A	C	NA	NA	NA	A	C	G
12	NA	A	C	NA	NA	NA	A	T	T
13	NA	A	C	NA	NA	NA	A	C	G
14	A	A	C	C	T	NA	A	T	T
15	A	A	C	T	G	C	A	T	G
16	A	A	C	T	T	C	C	C	G
17	A	A	C	T	T	C	A	T	G
22	NA	A	C	NA	NA	NA	C	C	G
23	A	T	C	C	T	NA	C	C	G
24	A	A	C	T	G	NA	A	C	G
25	NA	A	C	T	T	NA	A	C	G
26	A	A	C	T	T	NA	A	NA	G
27	A	A	C	C	G	C	A	C	G
28	A	T	C	C	T	C	A	T	T
29	A	T	C	C	T	C	A	C	G
31	NA	A	C	C	T	NA	C	C	G
33	A	A	C	T	T	NA	A	T	G
34	A	A	C	T	T	NA	A	C	G
35	A	A	C	NA	NA	NA	A	T	G
36	NA	A	T	C	NA	NA	A	NA	G
37	A	A	C	C	T	C	A	C	G
38	A	A	C	T	T	C	A	T	T
39	A	A	C	T	T	C	NA	NA	NA
40	NA	T	C	C	T	NA	A	T	G
41	NA	A	C	NA	NA	NA	A	C	G
42	NA	T	C	NA	NA	NA	A	T	G
43	NA	A	C	C	NA	NA	A	NA	G
44	NA	A	C	NA	NA	NA	A	NA	G
45	NA	A	C	C	T	NA	A	C	G
46	NA	A	C	NA	NA	NA	C	C	G
47	NA	A	C	NA	NA	NA	A	T	G
48	NA	A	C	NA	NA	NA	A	C	G
49	A	A	C	T	T	NA	A	T	G
50	NA	A	C	NA	NA	NA	A	T	G
51	NA	T	C	NA	NA	NA	A	NA	G
52	NA	A	C	NA	NA	NA	C	C	G
53	NA	A	C	NA	NA	NA	A	NA	G
54	A	A	C	C	T	NA	A	T	G
55	NA	A	T	C	T	NA	A	T	T
56	NA	A	C	NA	NA	NA	A	T	T
57	A	A	C	T	G	C	C	C	G
58	A	A	C	T	T	C	A	T	G
59	A	A	C	C	NA	C	A	C	G
60	NA	A	C	C	T	NA	A	T	G
61	A	NA	NA	NA	NA	C	NA	NA	NA
62	A	A	C	T	T	C	A	T	G
63	A	A	C	C	T	NA	A	T	G
64	A	A	C	C	T	C	A	T	G
65	NA	A	C	C	T	NA	A	NA	G
68	A	T	C	C	T	C	A	C	G
69	NA	A	T	C	T	C	A	C	G
70	A	A	C	T	G	C	C	C	G
71	A	A	T	C	T	C	A	T	G
73	A	A	C	T	T	C	A	C	G
74	A	A	T	C	T	C	A	C	G
75	A	A	C	C	G	C	A	C	G
76	NA	A	C	NA	NA	NA	A	T	G
79	NA	A	C	C	T	C	C	C	G
80	A	A	C	C	T	C	C	C	G
81	A	A	C	C	T	C	C	C	G
84	A	A	C	T	T	C	A	C	G
86	A	A	T	C	T	C	A	T	G
87	A	A	C	T	T	C	A	C	G
88	NA	A	C	NA	NA	NA	A	C	G
89	A	A	C	T	T	C	C	C	G
90	A	A	C	T	G	C	C	C	G
91	NA	A	C	C	T	C	A	T	T
92	A	A	C	C	T	C	C	C	G
93	A	A	NA	T	T	C	C	C	G
94	A	A	C	C	G	C	A	C	G
95	A	A	C	T	G	C	A	T	G
96	A	A	C	T	NA	C	A	C	G
97	A	A	C	T	T	C	A	C	G

Table A.3 (continued)

Line \ SNP	dome3UTR_(	dome5UTR_(	domeExon1_	domeExon1_	domeExon3_	domeExon3_	Dredd5inter_	DreddExon1_	DreddExon3_
98	NA	A	C	C	T	NA	A	C	G
99	NA	A	T	C	T	C	C	C	G
101	A	A	C	T	G	C	A	C	G
102	A	A	T	C	T	C	A	C	G
103	NA	A	C	C	T	NA	A	T	G
104	A	A	C	C	T	C	A	C	G
105	A	A	C	C	T	C	A	C	G
106	A	A	C	NA	T	C	A	T	G
107	A	A	C	T	G	C	C	C	G
108	NA	A	C	C	T	NA	A	C	G
109	A	A	NA	NA	NA	C	NA	NA	G
110	NA	A	C	T	T	C	NA	C	G
111	NA	A	C	C	NA	NA	A	C	G
112	NA	T	C	C	NA	NA	A	C	G
113	A	A	C	C	T	C	A	C	G
114	NA	A	C	C	T	C	C	C	G
115	A	A	C	T	T	C	A	C	G
116	A	A	C	T	G	C	A	T	G
117	NA	A	C	T	T	NA	C	C	G
119	A	A	C	T	T	C	A	C	G
122	A	A	C	T	T	C	C	C	G
123	NA	A	T	C	T	C	A	C	G
125	A	A	C	T	T	C	A	C	G
126	A	A	C	C	T	C	A	T	G
127	A	A	C	T	G	C	A	C	G
128	A	A	C	T	T	C	A	T	G
130	A	A	C	T	G	C	A	T	G
131	A	A	C	C	T	C	A	T	G
134	A	A	C	C	T	C	A	C	G
136	A	A	NA	C	T	C	A	C	G
137	NA	A	C	C	T	NA	NA	T	G
138	NA	A	T	C	T	C	A	C	G
139	A	A	C	C	T	C	A	T	NA
140	A	A	C	C	T	C	C	C	G
142	A	A	C	T	G	C	A	C	G
143	A	A	C	T	T	C	A	C	G
144	A	A	C	C	T	C	A	T	G
145	NA	A	C	C	T	NA	A	C	G
146	A	A	C	T	T	C	C	C	G
148	A	A	C	C	T	C	A	C	G
149	A	A	C	C	T	C	C	C	G
151	A	A	C	T	T	C	C	C	G
152	A	T	C	C	G	C	A	T	G
153	NA	T	C	C	T	NA	A	C	G
154	A	A	T	C	T	C	A	C	G
155	A	A	C	C	T	C	A	T	G
158	A	A	C	C	T	C	A	C	G
160	NA	A	T	C	T	C	C	C	G
164	NA	A	C	T	G	C	A	C	G
166	NA	A	C	C	T	C	A	T	G
167	A	A	C	C	NA	C	C	C	G
168	A	A	C	C	G	C	C	C	G
169	A	A	T	C	T	C	A	C	G
172	A	A	C	T	G	C	NA	T	G
173	A	A	C	C	G	C	A	C	G
174	A	A	C	T	T	C	A	T	T
201	A	A	C	T	G	C	A	C	G
202	A	A	C	T	G	C	A	C	G
203	A	A	C	C	T	C	A	C	G
204	A	A	C	C	T	C	A	NA	G
205	A	A	T	C	T	C	A	T	G
206	A	T	C	C	T	C	A	C	G
207	A	A	C	T	G	C	A	C	G
208	A	A	C	T	T	C	NA	C	G
209	A	A	C	T	T	C	A	C	G
210	A	A	C	T	T	C	C	C	G
211	NA	A	C	C	T	NA	NA	C	G
212	A	A	C	T	NA	C	NA	C	G
213	A	A	C	C	G	C	A	T	G
214	NA	A	C	C	T	NA	A	T	G
215	A	A	C	C	T	C	A	C	G
216	A	A	C	T	T	C	A	NA	G
217	A	A	C	T	T	C	A	C	G
218	A	A	C	T	T	C	A	T	T
219	NA	A	C	T	T	NA	C	C	G
220	A	A	C	NA	NA	C	C	C	G
221	NA	A	T	C	T	C	NA	NA	G
222	A	T	C	C	T	C	C	C	G
223	A	T	C	C	T	C	A	T	G
224	A	A	C	C	G	C	A	T	G
225	A	NA	C	C	T	C	A	T	G

Table A.3 (continued)

Line \ SNP	DreddIntron	Dsor15inter_	Dsor1Exon2_	hepExon1_0	hepExon2_0	hepExon3_0	hepExon4_0	hepExon5_0	hepExon5_0
1	A	NA	NA	T	NA	NA	G	A	NA
2	NA	NA	G	T	A	NA	G	NA	C
3	G	NA	NA	C	NA	NA	G	A	NA
4	G	NA	G	NA	A	NA	NA	NA	C
6	NA	T	G	T	A	A	G	A	C
7	G	NA	G	T	A	A	G	A	NA
8	NA	NA	G	T	A	NA	G	NA	C
9	G	C	G	T	A	NA	G	A	C
10	NA	NA	NA	T	NA	NA	G	A	NA
11	G	NA	NA	T	NA	NA	G	A	NA
12	G	NA	NA	T	NA	NA	G	A	NA
13	G	NA	NA	T	NA	NA	G	A	NA
14	G	NA	G	T	A	A	G	A	C
15	NA	NA	G	T	A	A	G	A	C
16	G	NA	G	T	A	A	G	A	C
17	G	C	G	T	A	T	T	G	C
22	G	NA	NA	T	NA	NA	T	G	NA
23	G	T	G	T	A	A	G	A	C
24	A	NA	A	C	NA	NA	T	G	C
25	NA	NA	G	T	A	NA	G	A	C
26	G	NA	A	T	A	A	G	A	C
27	G	C	G	T	A	A	G	A	NA
28	NA	C	G	T	A	A	G	A	C
29	NA	NA	G	T	A	A	G	A	C
31	NA	NA	G	T	A	A	G	A	C
33	G	C	G	T	A	A	G	A	C
34	G	NA	A	T	A	A	G	A	C
35	G	NA	A	T	A	A	G	A	C
36	NA	NA	G	T	A	NA	G	A	C
37	A	NA	G	T	A	A	G	A	C
38	G	NA	G	T	A	A	G	A	C
39	NA	NA	G	T	A	A	G	A	C
40	G	NA	G	T	A	NA	G	A	C
41	A	NA	NA	T	NA	NA	G	A	NA
42	G	NA	NA	T	NA	NA	T	G	NA
43	G	NA	NA	T	NA	NA	G	A	NA
44	NA	NA	NA	T	NA	NA	G	A	NA
45	A	NA	A	T	A	NA	G	A	C
46	NA	NA	NA	T	NA	NA	G	A	NA
47	G	T	NA	T	NA	NA	G	A	NA
48	NA	NA	NA	T	NA	NA	G	A	NA
49	G	NA	G	T	A	NA	G	A	NA
50	NA	NA	NA	T	NA	NA	G	A	NA
51	NA	NA	NA	T	NA	NA	NA	NA	NA
52	NA	NA	NA	T	NA	NA	G	A	NA
53	NA	NA	G	T	A	NA	G	A	C
54	NA	NA	G	T	A	NA	T	G	C
55	G	C	G	T	A	NA	G	A	C
56	NA	C	NA	T	NA	NA	G	A	NA
57	NA	NA	NA	NA	NA	NA	NA	NA	NA
58	G	C	G	T	A	A	G	A	C
59	NA	C	G	T	A	A	G	A	C
60	NA	NA	G	T	A	T	G	A	C
61	NA	NA	G	NA	G	T	NA	NA	C
62	NA	NA	G	T	A	A	G	A	A
63	NA	C	G	T	A	A	G	A	A
64	NA	NA	G	T	A	T	T	G	C
65	NA	NA	G	T	A	A	G	A	NA
68	NA	C	G	T	A	A	G	A	C
69	NA	NA	A	T	A	A	G	A	C
70	NA	T	G	C	G	T	T	G	C
71	G	NA	G	T	A	A	G	A	C
73	NA	C	G	T	A	A	G	A	C
74	NA	C	G	T	A	A	G	A	C
75	G	NA	A	T	A	A	G	A	C
76	NA	NA	NA	T	NA	NA	G	A	NA
79	NA	NA	G	T	A	A	G	A	C
80	NA	T	G	C	G	T	G	A	C
81	NA	T	G	T	A	A	G	A	C
84	A	T	G	T	A	A	G	A	A
86	NA	NA	A	T	A	A	G	A	C
87	NA	C	G	T	A	A	G	A	C
88	G	C	NA	T	A	NA	G	A	NA
89	NA	NA	G	T	A	A	G	A	C
90	G	T	G	T	A	A	G	A	C
91	NA	NA	G	C	A	T	G	A	C
92	NA	NA	G	T	A	A	G	A	C
93	G	NA	G	T	A	A	G	A	C
94	NA	C	G	T	A	A	G	A	A
95	G	NA	G	T	A	A	G	A	A
96	NA	NA	G	T	A	A	G	A	C
97	NA	NA	G	T	A	A	G	A	C

Table A.3 (continued)

Line \ SNP	DreddIntron	Dsor15inter_	Dsor1Exon2_	hepExon1_0	hepExon2_0	hepExon3_0	hepExon4_0	hepExon5_0	hepExon5_0
98	A	C	G	T	A	A	G	A	C
99	G	C	G	C	A	T	G	A	C
101	NA	C	G	T	A	A	G	A	C
102	NA	NA	G	T	A	A	G	A	C
103	G	C	G	T	A	A	G	A	C
104	NA	NA	G	T	A	A	G	A	C
105	NA	NA	G	T	A	A	G	A	C
106	G	NA	G	T	A	A	G	A	C
107	NA	NA	G	T	A	T	T	G	C
108	A	NA	G	T	A	A	G	A	C
109	G	NA	G	NA	A	T	NA	NA	C
110	NA	NA	G	T	A	A	G	A	C
111	G	T	NA	T	A	NA	T	G	NA
112	G	C	NA	T	A	NA	NA	A	NA
113	G	C	G	T	A	A	G	A	C
114	G	C	G	T	A	A	G	A	C
115	G	C	G	T	A	A	G	A	C
116	G	C	A	T	A	A	G	A	C
117	G	C	A	T	A	A	G	A	C
119	A	C	G	T	A	A	G	A	A
122	G	C	A	T	A	A	G	A	C
123	G	C	G	T	A	A	G	A	C
125	G	T	G	T	A	T	T	G	C
126	G	T	G	T	A	A	G	A	C
127	A	C	G	T	A	A	G	A	C
128	G	C	A	T	A	A	G	A	C
130	G	C	G	T	A	A	G	A	A
131	G	T	G	T	A	A	G	A	C
134	A	C	G	T	A	A	G	A	A
136	G	C	NA	T	A	A	G	A	NA
137	G	C	G	T	A	A	NA	A	C
138	A	C	G	T	A	A	G	A	C
139	G	C	G	T	A	A	G	A	C
140	G	C	G	T	A	A	G	A	C
142	G	C	NA	T	A	A	G	A	C
143	G	T	G	T	A	A	G	A	C
144	G	C	G	T	A	A	G	A	C
145	A	C	G	T	A	A	G	A	C
146	G	C	G	T	A	A	G	A	C
148	A	C	G	T	A	A	G	A	C
149	G	T	G	T	A	T	G	A	C
151	G	C	G	T	A	A	G	A	C
152	G	NA	G	T	A	A	G	A	C
153	G	NA	G	T	A	A	G	A	C
154	A	C	G	T	A	T	G	A	C
155	G	C	G	T	A	A	G	A	C
158	G	T	G	T	A	A	G	A	C
160	G	C	G	T	A	A	G	A	C
164	A	C	A	T	A	A	G	A	C
166	G	C	G	T	A	A	G	A	C
167	G	C	A	T	A	A	G	A	NA
168	G	C	A	T	A	A	G	A	A
169	G	C	G	T	A	A	G	A	C
172	G	C	G	T	A	A	G	A	C
173	G	C	G	T	A	A	G	A	C
174	G	C	G	T	A	T	G	A	C
201	A	C	G	T	A	A	G	A	C
202	G	T	G	T	A	A	G	A	C
203	A	T	G	C	A	T	G	G	C
204	NA	T	G	T	A	A	G	A	C
205	G	T	G	T	A	A	G	A	C
206	G	C	A	T	A	A	G	A	C
207	G	C	G	T	A	A	G	A	C
208	G	C	G	T	A	A	G	A	C
209	A	C	G	T	A	A	G	A	C
210	G	C	G	T	A	A	G	A	A
211	G	T	G	T	A	A	G	A	C
212	G	C	NA	T	A	A	G	A	C
213	G	C	G	T	A	A	G	A	C
214	G	C	G	T	A	A	G	A	C
215	A	C	G	T	A	A	G	A	C
216	G	C	G	T	A	A	G	A	A
217	G	C	G	T	A	A	G	A	C
218	G	C	G	T	A	T	T	G	C
219	G	NA	G	T	A	A	G	A	C
220	G	C	G	NA	NA	NA	T	G	C
221	G	C	G	T	A	A	G	A	C
222	G	C	G	T	A	A	G	A	C
223	G	NA	G	T	A	A	G	A	C
224	G	C	G	C	A	T	T	G	C
225	G	C	G	T	A	NA	G	A	C

Table A.3 (continued)

Line \ SNP	hepExon7_0	hepExon7_0	hepExon8_0	hepExon8_0	hepExon8_0	hepExon8_0	hepExon9_0	hepIntron1_	hopExon1_0
1	NA	T	NA	NA	NA	NA	NA	NA	NA
2	NA	NA	NA	NA	NA	NA	C	G	NA
3	T	T	NA	NA	NA	G	NA	NA	NA
4	NA	T	NA	NA	NA	G	C	NA	NA
6	C	T	NA	NA	A	G	C	G	NA
7	T	T	G	C	A	G	C	G	NA
8	NA	NA	NA	NA	NA	NA	C	G	NA
9	T	T	NA	NA	A	G	C	G	NA
10	T	T	NA	NA	NA	G	NA	NA	NA
11	T	T	NA	NA	NA	G	NA	NA	NA
12	T	C	NA	NA	NA	G	NA	NA	NA
13	T	T	NA	NA	NA	G	NA	NA	NA
14	C	T	NA	NA	G	A	C	G	NA
15	C	T	NA	C	A	G	C	G	NA
16	T	T	G	C	A	G	C	G	G
17	T	T	A	C	A	G	C	G	A
22	T	T	NA	NA	NA	G	NA	NA	NA
23	T	C	G	C	A	G	C	G	A
24	C	T	NA	NA	G	G	T	A	NA
25	T	T	NA	NA	A	G	C	G	NA
26	T	T	NA	C	A	G	C	G	A
27	T	T	G	C	A	G	C	G	NA
28	T	T	G	C	A	G	C	G	NA
29	T	T	G	C	A	G	C	G	A
31	T	NA	G	C	A	G	C	G	NA
33	C	T	NA	C	A	G	C	G	A
34	C	T	NA	NA	A	G	C	G	NA
35	T	T	NA	C	A	G	C	G	NA
36	T	T	NA	NA	NA	G	C	G	NA
37	T	C	G	C	A	G	C	G	A
38	T	T	NA	C	A	G	C	G	A
39	NA	T	G	C	A	G	C	G	NA
40	T	T	NA	NA	A	G	C	G	NA
41	T	C	NA	NA	NA	G	NA	NA	NA
42	T	T	NA	NA	NA	G	NA	NA	NA
43	T	T	NA	NA	NA	G	NA	NA	NA
44	T	T	NA	NA	NA	G	NA	NA	NA
45	NA	T	NA	NA	A	G	C	G	NA
46	T	T	NA	NA	NA	G	NA	NA	NA
47	C	T	NA	NA	NA	G	NA	NA	NA
48	T	T	NA	NA	NA	G	NA	NA	NA
49	T	T	NA	NA	A	G	C	G	NA
50	T	T	NA	NA	NA	G	NA	NA	NA
51	T	T	NA	NA	NA	G	NA	NA	NA
52	T	T	NA	NA	NA	G	NA	NA	NA
53	T	T	NA	NA	A	G	C	G	NA
54	C	T	NA	NA	A	A	C	G	NA
55	T	T	NA	NA	A	G	C	G	NA
56	T	T	NA	NA	NA	G	NA	NA	NA
57	NA	T	G	NA	NA	G	NA	NA	NA
58	T	T	G	C	A	G	C	G	A
59	T	T	G	C	A	G	C	G	A
60	T	T	G	C	A	G	C	G	A
61	NA	NA	G	NA	G	NA	T	A	A
62	T	T	G	C	A	G	C	G	A
63	T	T	G	NA	A	G	C	G	A
64	T	T	A	C	A	G	C	G	A
65	T	T	G	C	A	G	C	G	NA
68	C	T	G	C	A	G	C	G	A
69	T	T	G	C	A	G	C	G	A
70	C	T	G	A	G	G	T	A	G
71	NA	T	G	C	A	G	C	G	A
73	T	C	G	C	A	G	C	G	A
74	T	T	G	C	A	G	C	G	NA
75	T	T	G	C	A	G	C	G	NA
76	T	T	NA	NA	NA	G	NA	NA	NA
79	T	T	G	C	A	G	C	G	A
80	T	T	G	C	A	G	C	A	G
81	T	T	G	C	A	G	C	G	G
84	T	T	G	C	A	G	C	G	G
86	T	T	G	C	A	G	C	G	A
87	T	T	G	C	NA	NA	C	G	G
88	T	T	NA	NA	NA	G	NA	G	NA
89	NA	T	G	NA	NA	G	NA	G	A
90	C	T	G	C	A	G	C	G	NA
91	T	C	G	C	A	G	C	A	A
92	T	C	G	C	A	G	C	G	A
93	T	T	G	C	A	G	C	G	A
94	T	T	G	C	A	G	C	G	A
95	T	T	G	C	A	G	C	G	A
96	T	T	G	C	A	G	C	G	A
97	T	T	G	C	A	G	C	G	A

Table A.3 (continued)

Line \ SNP	hepExon7_0	hepExon7_0	hepExon8_0	hepExon8_0	hepExon8_0	hepExon8_0	hepExon9_0	hepIntron1_	hopExon1_0
98	NA	T	A	C	A	G	C	G	A
99	T	T	G	C	A	G	C	A	A
101	T	T	G	C	A	G	C	G	G
102	T	T	G	C	A	G	C	G	G
103	T	T	G	C	A	G	C	G	A
104	T	T	G	C	A	G	C	G	A
105	T	C	G	C	A	G	C	G	A
106	T	T	G	C	A	G	C	G	NA
107	NA	T	A	C	A	A	C	G	A
108	C	T	A	C	A	G	C	G	NA
109	NA	T	G	C	NA	G	C	G	NA
110	T	T	G	C	A	G	C	G	A
111	C	T	NA	NA	NA	A	C	G	NA
112	C	NA	G	NA	NA	NA	C	G	NA
113	C	T	A	C	A	G	C	G	A
114	C	T	A	C	A	G	C	G	A
115	C	T	G	C	A	G	C	G	G
116	T	NA	G	C	A	G	C	G	G
117	T	T	G	C	A	G	C	G	A
119	T	NA	G	C	A	G	C	G	A
122	T	T	G	C	A	G	C	G	A
123	T	T	G	C	A	G	C	G	A
125	T	T	A	C	A	G	C	G	A
126	T	T	G	C	A	G	C	G	G
127	T	T	G	C	A	G	C	G	A
128	T	T	G	C	A	G	C	G	A
130	T	NA	G	C	A	G	C	G	A
131	C	T	G	C	A	G	C	G	A
134	T	T	G	C	A	G	C	G	A
136	T	T	G	C	A	G	C	G	A
137	T	NA	G	C	A	G	C	G	A
138	T	T	G	C	A	G	C	G	A
139	T	T	G	C	A	G	C	G	A
140	T	T	G	C	A	G	C	G	A
142	T	T	G	C	A	G	C	G	NA
143	C	T	G	C	A	G	C	G	G
144	C	T	G	C	G	A	C	G	G
145	T	T	G	C	A	G	C	G	G
146	T	T	G	C	A	G	C	G	A
148	T	T	G	C	A	G	C	G	A
149	T	T	G	C	A	G	C	G	A
151	T	T	G	C	A	G	C	G	A
152	T	T	G	C	A	G	C	G	A
153	T	C	G	C	A	G	C	G	A
154	T	T	G	C	A	G	C	G	A
155	NA	T	G	C	A	G	C	G	A
158	NA	T	G	C	A	G	C	G	G
160	T	T	G	C	A	G	C	G	A
164	T	T	G	C	A	G	C	G	A
166	T	T	G	C	A	G	C	G	G
167	T	NA	G	C	A	G	C	G	A
168	T	T	G	C	A	G	C	G	A
169	C	T	A	C	A	G	C	G	A
172	T	T	G	C	A	G	C	G	A
173	T	T	G	C	A	G	C	G	A
174	T	T	G	C	A	G	C	G	G
201	C	T	G	C	A	G	C	G	A
202	T	T	G	C	A	G	C	G	A
203	C	T	A	A	G	G	C	G	A
204	T	T	G	C	A	G	C	G	NA
205	T	T	G	C	A	G	C	G	A
206	T	C	G	C	A	G	C	G	A
207	C	T	G	C	A	G	C	G	G
208	T	C	G	C	A	G	C	G	A
209	T	T	G	C	A	G	C	G	A
210	T	T	G	C	A	G	C	G	A
211	C	T	G	C	A	G	C	G	G
212	C	T	G	C	A	G	C	G	NA
213	T	T	G	C	A	G	C	G	A
214	T	T	G	C	A	G	C	G	A
215	T	T	G	C	A	G	C	G	G
216	T	T	G	C	A	G	C	G	G
217	T	T	G	C	A	G	C	G	G
218	T	T	A	C	A	G	C	G	A
219	T	NA	G	C	A	G	C	G	A
220	NA	T	G	C	A	NA	C	NA	NA
221	T	T	G	C	A	G	C	G	NA
222	T	T	G	C	A	G	C	G	G
223	T	T	G	C	A	G	C	G	A
224	C	T	G	C	G	G	C	A	A
225	NA	T	NA	C	A	G	C	G	A

Table A.3 (continued)

Line \ SNP	hopExon1_0	hopExon4_0	hopExon4_0	hopExon7_0	hopExon7_0	hopExon8_0	lzExon5_01	lzExon5_02	lzExon5_03
1	G	NA	NA	NA	C	NA	NA	NA	NA
2	G	NA	NA	NA	C	NA	T	NA	NA
3	G	NA	G	NA	C	G	NA	NA	NA
4	G	NA	G	NA	C	NA	T	G	T
6	G	G	G	NA	C	G	NA	NA	C
7	G	NA	G	A	C	NA	C	NA	NA
8	G	NA	NA	NA	C	NA	NA	NA	NA
9	G	NA	G	A	C	NA	T	NA	C
10	G	NA	G	NA	C	G	NA	NA	NA
11	G	NA	G	NA	C	G	NA	G	NA
12	G	NA	C	NA	C	G	NA	NA	NA
13	G	NA	G	NA	C	T	NA	NA	NA
14	G	NA	G	A	C	NA	NA	NA	NA
15	G	G	C	A	C	T	NA	T	T
16	A	G	G	A	C	G	C	G	C
17	A	G	G	A	C	G	T	G	T
22	A	NA	G	NA	C	G	NA	NA	NA
23	G	G	G	A	C	G	T	NA	T
24	G	NA	G	G	C	NA	NA	T	C
25	G	G	G	G	C	NA	T	NA	NA
26	G	G	G	A	C	T	NA	G	C
27	G	G	G	A	C	G	C	NA	C
28	A	G	G	A	C	NA	T	NA	T
29	G	A	C	A	C	G	C	NA	C
31	G	G	G	A	C	G	C	NA	T
33	G	G	G	A	C	T	NA	NA	C
34	A	G	G	NA	C	G	NA	T	C
35	G	G	C	A	C	T	T	NA	T
36	NA	NA	G	NA	C	NA	NA	NA	NA
37	G	G	G	A	C	T	C	G	T
38	A	G	G	A	C	T	C	NA	T
39	NA	G	G	A	C	NA	C	NA	C
40	NA	NA	G	NA	C	G	NA	G	C
41	G	NA	G	NA	C	NA	NA	NA	NA
42	A	NA	G	NA	C	NA	NA	NA	NA
43	G	NA	G	NA	C	T	NA	NA	NA
44	G	NA	G	NA	C	NA	NA	NA	NA
45	G	NA	G	NA	C	G	NA	NA	NA
46	G	NA	G	NA	C	T	NA	NA	NA
47	G	NA	C	NA	C	G	NA	G	NA
48	NA	NA	G	NA	C	G	NA	NA	NA
49	G	G	G	NA	C	T	NA	G	C
50	G	NA	G	NA	C	T	NA	G	NA
51	NA	NA	G	NA	C	T	NA	NA	NA
52	G	NA	G	NA	C	G	NA	G	NA
53	NA	NA	NA	NA	C	G	NA	NA	C
54	G	G	G	A	C	G	T	G	T
55	G	NA	G	NA	C	G	NA	NA	C
56	G	NA	C	NA	C	T	NA	NA	NA
57	G	G	G	NA	C	G	C	NA	C
58	A	G	G	A	C	T	T	NA	C
59	G	G	G	A	C	G	T	NA	T
60	A	G	G	A	C	T	T	NA	NA
61	NA	G	NA	A	NA	NA	C	NA	C
62	G	G	G	A	C	T	T	G	T
63	G	G	G	A	C	T	C	G	NA
64	G	G	G	A	C	G	C	NA	C
65	G	G	G	NA	C	G	NA	NA	C
68	G	G	G	A	C	G	C	G	NA
69	G	G	G	A	C	G	C	G	C
70	G	G	G	G	C	NA	C	G	C
71	A	G	G	A	C	G	NA	NA	NA
73	G	G	G	A	C	T	T	G	T
74	A	G	G	A	C	G	C	NA	NA
75	A	A	C	A	C	G	T	NA	T
76	G	NA	G	NA	C	T	NA	NA	NA
79	G	G	G	A	C	NA	C	NA	C
80	G	G	C	A	C	G	C	NA	C
81	A	G	G	A	C	G	C	G	NA
84	A	G	G	A	C	G	C	NA	C
86	A	G	G	A	C	G	C	G	NA
87	G	G	G	G	C	G	C	G	NA
88	G	NA	G	A	C	NA	NA	NA	NA
89	G	G	G	A	C	NA	T	NA	C
90	G	A	G	A	C	NA	NA	NA	T
91	G	G	G	A	C	T	C	NA	C
92	G	G	G	A	C	G	C	G	C
93	A	G	G	A	C	G	T	G	C
94	G	G	G	A	C	G	NA	NA	NA
95	G	G	G	A	C	G	C	NA	T
96	G	G	G	A	C	G	T	NA	T
97	G	G	G	A	C	G	NA	G	NA

Table A.3 (continued)

Line \ SNP	hopExon1_01	hopExon4_01	hopExon4_02	hopExon7_01	hopExon7_02	hopExon8_01	lzExon5_01	lzExon5_02	lzExon5_03
98	A	G	G	A	C	G	C	NA	C
99	G	G	G	A	C	G	NA	NA	T
101	G	G	C	A	C	G	T	NA	T
102	G	G	G	A	C	NA	NA	NA	C
103	A	G	G	A	T	G	C	NA	T
104	A	G	G	A	C	G	C	NA	C
105	G	G	G	A	C	T	T	NA	T
106	G	G	G	A	C	G	NA	NA	NA
107	G	G	G	A	C	G	T	NA	T
108	NA	NA	G	A	C	T	NA	NA	NA
109	G	G	NA	A	C	NA	NA	NA	NA
110	G	G	G	A	C	T	T	G	T
111	G	NA	G	NA	C	NA	NA	NA	T
112	G	NA	G	NA	C	NA	T	NA	T
113	G	G	G	A	C	T	C	G	C
114	A	G	G	A	C	G	T	G	C
115	G	G	G	A	C	G	T	NA	C
116	G	G	C	A	C	NA	T	NA	T
117	G	G	G	A	C	G	C	G	C
119	G	G	G	A	C	NA	T	NA	T
122	G	G	G	A	C	G	C	G	C
123	G	A	G	A	C	T	C	G	T
125	G	G	G	A	C	G	C	G	C
126	A	G	G	A	C	T	C	G	T
127	G	G	G	A	C	G	C	G	C
128	G	G	G	A	T	G	T	G	T
130	G	A	G	A	C	G	C	NA	C
131	G	A	G	A	C	T	C	G	C
134	G	G	G	A	C	G	C	G	C
136	G	G	G	A	C	NA	C	NA	C
137	G	A	G	A	C	T	C	NA	C
138	G	G	G	A	C	G	C	G	C
139	A	G	G	A	C	G	C	G	C
140	G	G	G	A	C	T	T	G	T
142	G	G	G	A	NA	G	T	G	T
143	G	G	G	A	C	G	T	G	T
144	A	G	G	A	C	G	T	G	T
145	G	G	C	A	C	NA	C	NA	C
146	A	A	C	A	C	G	C	G	T
148	G	G	G	A	C	G	C	G	T
149	G	G	G	A	C	G	T	G	T
151	G	G	G	A	C	G	C	G	C
152	G	G	G	A	C	G	T	G	T
153	G	A	G	A	C	G	T	NA	T
154	G	A	G	A	C	G	C	NA	C
155	G	A	G	A	C	G	C	G	C
158	G	G	G	A	C	G	C	NA	C
160	G	G	G	A	C	G	C	NA	C
164	G	A	G	A	C	G	C	G	C
166	G	G	G	A	C	G	C	T	C
167	G	G	G	A	C	NA	C	NA	C
168	G	G	G	A	C	NA	C	NA	C
169	A	G	G	A	C	G	C	G	C
172	G	G	G	A	C	G	T	G	T
173	G	G	G	A	C	G	C	NA	C
174	G	G	G	G	C	G	C	NA	C
201	G	G	C	A	C	G	T	NA	T
202	A	G	G	A	C	G	T	NA	C
203	G	G	G	A	C	NA	C	NA	T
204	G	G	G	NA	C	NA	C	G	NA
205	G	G	G	A	C	T	C	G	T
206	G	G	G	A	C	G	T	G	C
207	G	G	C	A	C	G	C	G	C
208	A	G	G	A	C	NA	T	NA	T
209	G	G	G	A	C	NA	C	NA	T
210	G	G	G	A	C	T	C	T	C
211	G	G	G	A	C	NA	C	NA	C
212	G	NA	NA	A	C	G	NA	NA	NA
213	G	G	G	A	C	G	C	NA	C
214	G	G	G	A	C	G	C	G	C
215	G	G	G	A	C	T	C	G	C
216	G	G	G	G	C	NA	C	NA	C
217	A	G	G	A	C	G	T	NA	C
218	G	G	G	A	C	G	T	G	T
219	G	A	G	A	C	NA	T	NA	T
220	NA	G	NA	A	C	G	NA	NA	NA
221	NA	G	NA	A	NA	G	T	NA	NA
222	G	G	G	A	C	NA	T	NA	T
223	G	G	G	A	C	G	T	NA	T
224	G	G	G	A	C	G	C	T	C
225	G	G	G	A	C	NA	C	NA	C

Table A.3 (continued)

Line \ SNP	lzIntron1_02	mxc5inter_0	mxcExon2_0	Ntf25inter_0	Ntf2Intron1_	osExon1_01	PGRPLE3inte	phl3UTR_01	phl5UTR_01
1	NA	C	NA	NA	NA	C	C	NA	NA
2	NA	NA	A	NA	NA	C	NA	A	NA
3	NA	A	NA	NA	NA	C	G	NA	NA
4	NA	NA	NA	NA	NA	C	C	A	NA
6	A	C	A	C	G	C	G	NA	NA
7	G	C	A	G	NA	C	C	G	A
8	NA	NA	NA	NA	NA	C	C	NA	NA
9	A	NA	T	G	NA	C	C	NA	A
10	NA	C	NA	NA	NA	C	C	NA	NA
11	NA	C	NA	NA	NA	C	C	NA	NA
12	NA	A	NA	NA	NA	C	C	NA	NA
13	NA	A	NA	NA	NA	C	G	NA	NA
14	NA	A	T	C	G	C	C	G	A
15	G	A	T	G	G	C	C	A	A
16	G	A	T	G	G	C	C	A	A
17	G	A	T	G	A	C	G	G	G
22	NA	A	NA	NA	NA	C	C	NA	NA
23	G	C	A	G	G	A	C	A	G
24	A	A	T	NA	A	C	G	A	G
25	G	A	T	C	G	C	G	A	A
26	A	A	T	C	G	C	C	G	NA
27	G	C	A	G	G	C	C	G	A
28	G	C	A	G	G	C	C	G	A
29	G	C	A	G	A	C	C	G	A
31	G	C	A	G	G	C	C	A	G
33	A	A	A	G	A	C	C	A	A
34	A	A	T	C	G	C	C	A	G
35	A	A	T	G	NA	C	C	A	A
36	NA	A	NA	NA	NA	C	NA	NA	NA
37	G	NA	NA	G	G	C	C	G	A
38	G	C	A	G	G	C	C	G	A
39	G	NA	NA	G	G	C	NA	NA	NA
40	NA	C	A	C	NA	C	C	G	A
41	NA	A	NA	NA	NA	C	G	NA	NA
42	NA	C	NA	NA	NA	C	C	NA	NA
43	NA	C	NA	NA	NA	C	C	NA	NA
44	NA	NA	NA	NA	NA	C	C	NA	NA
45	NA	A	T	NA	NA	C	C	NA	NA
46	NA	A	NA	NA	NA	C	C	NA	NA
47	NA	A	NA	NA	NA	C	C	NA	NA
48	NA	A	NA	NA	NA	C	NA	NA	NA
49	G	A	A	NA	NA	C	G	A	G
50	NA	A	NA	NA	NA	C	G	NA	NA
51	NA	A	NA	NA	NA	C	C	NA	NA
52	NA	C	NA	NA	NA	C	C	NA	NA
53	NA	A	T	NA	NA	C	NA	G	NA
54	NA	C	NA	G	A	C	C	G	A
55	NA	A	T	NA	NA	C	C	A	NA
56	NA	A	NA	NA	NA	C	C	NA	NA
57	NA	A	T	G	G	C	NA	A	A
58	A	A	T	G	G	C	G	G	G
59	G	C	A	C	G	C	C	G	A
60	G	C	A	G	G	C	NA	A	A
61	G	NA	T	G	NA	NA	NA	A	A
62	G	C	A	G	G	C	NA	A	G
63	G	C	A	C	G	C	NA	A	A
64	A	A	A	C	G	C	C	A	G
65	A	NA	A	G	NA	C	NA	A	A
68	G	C	A	C	A	C	G	A	A
69	G	A	T	C	G	C	C	A	G
70	G	C	A	G	G	C	G	G	A
71	G	C	A	G	A	C	C	A	A
73	G	C	A	G	G	C	G	A	G
74	G	A	A	G	G	C	G	A	A
75	A	A	T	G	A	C	C	A	A
76	NA	A	NA	NA	NA	C	G	NA	NA
79	G	NA	A	C	G	C	C	A	A
80	A	C	A	C	G	C	G	G	A
81	G	C	A	C	G	C	NA	A	A
84	G	A	A	G	A	C	C	G	A
86	G	A	T	G	A	C	C	A	A
87	G	C	A	G	G	C	C	A	A
88	NA	NA	NA	NA	NA	C	C	NA	NA
89	A	A	T	G	G	C	G	A	A
90	A	C	A	C	G	C	C	G	A
91	G	C	A	G	G	C	G	G	A
92	G	A	A	C	G	C	C	G	G
93	G	A	T	C	G	C	C	A	A
94	G	A	T	G	A	C	C	A	A
95	G	A	A	G	G	C	C	G	A
96	G	A	T	G	G	C	G	G	A
97	G	C	A	C	G	C	G	A	G

Table A.3 (continued)

Line \ SNP	lzIntron1_02	mxc5inter_0	mxcExon2_0	Ntf25inter_0	Ntf2Intron1_	osExon1_01	PGRPLE3inte	phl3UTR_01	phl5UTR_01
98	G	A	A	G	G	C	C	A	G
99	A	A	T	G	G	C	C	A	G
101	A	C	A	G	A	C	C	A	A
102	A	A	T	G	A	C	C	NA	NA
103	G	A	A	C	G	C	C	A	A
104	G	C	A	G	G	C	C	A	G
105	G	A	T	G	A	C	G	A	A
106	A	NA	A	NA	NA	C	NA	A	G
107	G	NA	A	G	G	C	G	A	A
108	NA	A	T	G	G	C	C	G	NA
109	NA	NA	NA	NA	G	C	NA	G	A
110	G	C	A	C	G	C	C	A	NA
111	A	C	A	NA	NA	C	C	NA	NA
112	NA	C	A	NA	NA	C	G	G	A
113	G	A	T	C	G	C	G	A	A
114	G	A	NA	G	G	C	C	A	A
115	G	A	T	C	G	C	C	G	A
116	A	A	T	NA	G	C	G	G	G
117	A	A	T	G	G	C	G	G	A
119	A	NA	A	G	G	C	NA	A	A
122	G	A	T	C	G	C	C	A	G
123	A	C	A	G	G	C	C	A	A
125	G	A	A	C	G	C	G	A	G
126	G	C	A	C	G	C	C	G	A
127	A	C	A	G	A	C	G	G	G
128	A	A	T	G	G	C	C	A	A
130	G	A	T	C	G	C	C	G	A
131	G	C	A	G	A	A	C	A	A
134	A	C	A	C	G	C	G	A	A
136	A	NA	NA	NA	G	C	NA	A	G
137	A	NA	A	C	G	C	C	G	A
138	G	C	A	C	G	C	C	A	G
139	G	A	A	C	G	C	C	NA	NA
140	G	C	A	C	G	C	C	G	A
142	NA	NA	NA	G	G	C	C	NA	NA
143	A	C	A	G	G	C	C	A	A
144	A	A	T	G	G	C	C	A	A
145	G	NA	A	C	G	C	C	A	G
146	A	A	A	G	A	C	C	A	A
148	G	C	A	G	G	C	C	A	G
149	G	C	A	G	A	C	C	A	A
151	A	C	A	C	G	C	G	A	G
152	G	A	T	G	A	C	C	A	G
153	G	A	T	C	G	C	C	A	A
154	G	A	T	G	A	C	C	A	A
155	G	A	T	G	A	C	C	A	A
158	G	C	A	C	G	C	G	A	G
160	G	A	T	G	G	C	G	A	A
164	A	A	T	C	G	C	C	A	A
166	G	C	A	C	G	C	C	G	A
167	G	A	T	NA	A	C	NA	A	G
168	G	A	T	C	G	C	G	A	G
169	G	C	A	G	A	C	G	A	A
172	G	A	T	C	G	C	G	A	A
173	G	A	A	NA	G	C	C	G	G
174	G	C	A	G	G	C	C	G	A
201	A	C	A	G	G	C	G	A	A
202	G	C	A	G	A	C	G	G	A
203	A	C	A	G	G	C	G	A	A
204	NA	C	A	G	NA	C	C	A	A
205	A	C	A	G	A	C	C	A	A
206	G	A	T	G	A	C	C	G	A
207	A	A	A	G	G	C	G	A	A
208	G	A	T	NA	G	C	G	A	A
209	A	A	T	G	A	C	G	A	A
210	G	C	A	G	A	C	C	A	A
211	G	C	A	G	G	C	C	A	A
212	NA	A	T	G	NA	C	C	NA	A
213	G	A	T	G	G	C	C	A	G
214	G	C	A	G	A	C	G	A	A
215	G	C	A	C	G	C	G	A	A
216	A	C	A	G	G	C	NA	G	NA
217	G	A	A	G	A	C	C	A	G
218	G	A	T	G	A	A	C	G	A
219	G	A	T	G	A	C	G	G	A
220	G	A	NA	G	NA	C	G	A	G
221	G	A	NA	G	G	C	G	NA	NA
222	G	A	T	G	G	C	G	A	G
223	G	A	T	G	G	C	C	A	A
224	A	A	T	G	G	C	G	G	A
225	A	A	A	C	G	C	NA	A	A

Table A.3 (continued)

Line \ SNP	phl5UTR_02	phlExon2_01	phlExon2_02	phlExon2_03	phlExon4_01	phlExon4_02	phlExon5_01	Pvf15UTR_0	Pvf15UTR_0
1	NA	NA	NA	NA	NA	NA	NA	T	C
2	G	C	G	NA	T	NA	G	NA	NA
3	NA	NA	NA	NA	NA	NA	NA	A	C
4	G	C	G	G	T	T	A	A	C
6	G	NA	G	NA	T	NA	A	A	C
7	G	C	G	G	T	NA	A	A	NA
8	G	NA	NA	NA	NA	NA	G	NA	NA
9	G	T	A	G	C	NA	G	A	C
10	NA	NA	NA	NA	NA	NA	NA	A	T
11	NA	NA	NA	NA	NA	NA	NA	A	T
12	NA	NA	NA	NA	NA	NA	NA	A	T
13	NA	NA	NA	NA	NA	NA	NA	A	NA
14	A	C	G	G	T	T	A	NA	NA
15	G	T	A	G	C	C	G	NA	C
16	G	C	G	G	T	T	A	A	T
17	G	T	G	G	T	C	G	A	T
22	NA	NA	NA	NA	NA	NA	NA	A	C
23	G	C	G	G	T	T	A	A	C
24	G	T	G	NA	T	T	A	A	T
25	G	T	A	G	C	C	G	A	C
26	NA	C	G	G	T	NA	A	T	C
27	G	C	G	G	T	T	A	A	T
28	G	C	G	G	T	T	A	A	T
29	G	T	G	A	T	C	G	A	T
31	G	T	G	A	T	C	A	A	C
33	G	T	A	G	C	C	G	A	C
34	G	C	G	G	T	T	A	T	C
35	G	C	G	G	T	T	A	NA	C
36	G	T	G	NA	T	NA	A	A	C
37	G	C	G	G	T	T	A	A	NA
38	A	C	G	G	T	T	A	A	C
39	G	T	NA	NA	NA	C	G	NA	C
40	G	C	G	G	T	T	A	A	T
41	NA	NA	NA	NA	NA	NA	NA	A	C
42	NA	NA	NA	NA	NA	NA	NA	A	T
43	NA	NA	G	NA	T	NA	A	A	C
44	NA	NA	NA	NA	NA	NA	NA	A	NA
45	G	T	A	G	C	NA	G	A	T
46	NA	T	NA	NA	NA	NA	NA	A	T
47	NA	T	A	NA	C	NA	NA	A	T
48	NA	NA	NA	NA	NA	NA	NA	A	NA
49	G	C	G	G	T	T	A	A	C
50	NA	NA	NA	NA	NA	NA	NA	T	C
51	NA	NA	NA	NA	NA	NA	NA	NA	NA
52	NA	NA	NA	NA	NA	NA	NA	A	C
53	G	NA	G	NA	T	T	A	A	C
54	G	C	G	G	T	T	A	A	C
55	G	T	G	NA	T	C	A	NA	C
56	NA	NA	NA	NA	T	NA	NA	T	C
57	G	C	G	G	T	T	A	NA	NA
58	G	T	G	A	T	C	G	NA	C
59	G	T	G	A	T	C	G	NA	NA
60	G	T	A	G	C	C	G	A	C
61	G	C	G	G	T	T	A	NA	NA
62	G	C	G	G	T	T	NA	A	C
63	G	T	A	G	C	C	G	NA	C
64	G	C	G	G	T	T	A	A	C
65	G	NA	NA	G	NA	NA	NA	A	C
68	G	T	A	G	T	C	G	A	C
69	G	C	G	G	T	T	A	A	T
70	G	C	G	G	T	T	A	NA	T
71	G	T	A	G	C	C	G	A	C
73	G	C	G	G	T	T	A	A	T
74	G	C	G	G	T	NA	A	NA	C
75	G	C	G	G	T	C	G	A	T
76	NA	NA	NA	NA	NA	NA	NA	A	C
79	G	C	G	G	T	T	A	A	C
80	G	C	G	G	T	T	A	A	C
81	G	C	G	G	T	T	A	A	T
84	G	C	G	G	T	T	A	A	C
86	G	C	G	G	T	T	A	A	T
87	G	C	G	G	T	T	A	NA	C
88	G	C	NA	NA	T	T	NA	NA	NA
89	G	T	A	G	C	C	G	A	C
90	G	C	G	NA	T	T	A	A	C
91	A	C	G	G	T	T	A	A	C
92	G	T	G	A	T	C	G	A	C
93	G	T	A	G	C	C	G	A	C
94	G	C	G	G	T	T	A	A	C
95	G	C	G	G	T	T	A	A	C
96	G	T	A	G	C	C	G	A	NA
97	G	T	G	A	T	C	A	A	T

Table A.3 (continued)

Line \ SNP	phl5UTR_02	phlExon2_01	phlExon2_02	phlExon2_03	phlExon4_01	phlExon4_02	phlExon5_01	Pvf15UTR_0	Pvf15UTR_0
98	G	T	G	A	T	T	A	A	C
99	G	C	G	G	T	NA	A	A	C
101	G	T	A	G	C	C	G	T	C
102	NA	C	G	G	T	T	A	NA	C
103	G	C	G	G	T	T	A	A	C
104	G	C	G	G	T	T	A	NA	T
105	G	C	G	G	T	T	A	A	T
106	G	C	G	G	T	T	A	A	C
107	G	C	G	G	T	T	A	NA	T
108	G	T	G	G	T	NA	G	A	C
109	G	C	G	G	T	NA	NA	NA	T
110	G	C	G	G	T	T	A	A	NA
111	G	C	G	G	T	T	A	A	C
112	G	T	NA	G	C	C	NA	A	C
113	G	T	A	G	C	C	G	A	C
114	G	C	G	G	T	T	A	A	C
115	A	C	G	G	T	T	A	A	C
116	G	T	G	G	T	C	G	A	C
117	G	C	G	G	T	T	A	A	T
119	G	C	G	G	T	T	A	A	NA
122	G	C	G	G	T	T	A	A	C
123	G	C	G	G	T	T	A	A	C
125	G	C	G	G	T	T	A	A	C
126	G	C	G	G	T	T	A	A	T
127	G	T	G	A	T	C	G	A	NA
128	G	T	A	G	C	C	G	A	T
130	G	C	G	G	T	T	A	A	NA
131	G	C	G	G	C	C	G	A	T
134	G	T	A	G	C	C	G	A	T
136	G	C	G	G	T	T	A	A	T
137	G	C	G	G	T	T	A	A	C
138	G	T	G	A	T	T	A	A	T
139	G	C	G	G	T	T	A	A	C
140	G	C	G	G	T	T	A	A	C
142	NA	C	G	G	T	T	A	A	C
143	G	C	G	G	T	T	A	A	C
144	G	C	G	G	C	C	G	A	C
145	G	C	G	G	T	T	A	A	C
146	G	T	A	G	C	C	G	A	T
148	G	T	G	A	T	T	A	A	C
149	G	C	G	G	T	T	A	A	C
151	G	T	G	A	T	C	A	A	C
152	G	C	G	G	T	T	A	A	C
153	G	T	A	G	C	C	G	A	C
154	G	T	A	G	C	C	G	A	T
155	G	C	G	G	T	T	A	A	T
158	G	C	G	G	T	T	A	A	NA
160	G	C	G	G	T	T	A	A	T
164	G	T	A	G	C	C	G	A	C
166	G	C	G	G	T	T	A	A	C
167	G	C	G	G	T	T	A	A	NA
168	G	C	G	G	T	C	A	T	C
169	G	C	G	G	T	T	A	A	T
172	G	T	A	G	C	C	G	A	T
173	G	T	G	G	T	T	G	A	NA
174	A	C	G	G	T	T	NA	A	NA
201	G	C	G	G	T	T	A	A	T
202	G	C	G	G	T	T	A	A	T
203	G	C	G	G	C	C	G	A	T
204	G	NA	G	NA	T	NA	NA	A	C
205	G	C	G	G	T	T	A	A	C
206	G	C	G	G	T	T	A	A	C
207	G	C	G	G	T	T	A	A	C
208	G	T	NA	G	NA	NA	NA	A	C
209	G	C	G	G	C	C	G	A	C
210	G	C	G	G	T	T	A	A	C
211	G	C	G	G	T	T	A	A	C
212	G	C	G	G	T	T	A	A	C
213	G	C	G	G	T	T	A	A	T
214	G	C	G	G	T	T	A	A	T
215	G	T	G	A	T	C	G	A	C
216	G	T	G	G	T	C	NA	A	C
217	G	C	G	G	T	T	A	A	C
218	A	C	G	G	T	T	A	A	T
219	G	C	G	G	T	T	G	A	T
220	G	NA	G	NA	T	NA	A	A	T
221	G	NA	G	NA	T	NA	A	A	C
222	G	C	G	G	T	T	A	A	T
223	G	C	G	G	T	T	A	A	T
224	G	C	G	G	T	T	A	A	C
225	G	T	A	G	C	C	G	A	C

Table A.3 (continued)

Line \ SNP	Pvf1Exon1_C	Pvf1Exon1_C	Pvf1Exon4_C	Pvf1Exon4_C	Pvf1Exon5_C	Pvf1Intron1_	Pvf1Intron1_	Pvf1Intron4_	Rps65inter_C
1	G	NA	NA	C	NA	NA	NA	NA	NA
2	NA	NA	NA	NA	NA	NA	A	A	A
3	G	NA	NA	C	C	G	A	NA	NA
4	NA	NA	NA	C	C	NA	A	A	NA
6	T	NA	NA	C	NA	G	C	A	NA
7	NA	NA	A	C	C	G	NA	A	NA
8	NA	NA	NA	C	NA	NA	A	A	NA
9	T	NA	A	C	NA	G	A	A	A
10	G	NA	NA	C	C	G	A	NA	NA
11	G	G	NA	C	C	A	C	NA	NA
12	T	NA	NA	C	C	G	A	NA	NA
13	T	A	NA	C	C	G	A	NA	NA
14	G	NA	A	C	C	NA	NA	A	A
15	T	NA	A	C	C	G	C	A	G
16	T	A	G	C	C	G	A	A	G
17	G	G	A	C	C	A	A	A	A
22	G	G	NA	C	C	G	A	NA	NA
23	G	G	G	C	C	G	A	A	G
24	G	NA	NA	C	C	A	NA	A	A
25	T	NA	NA	C	C	G	A	A	A
26	G	NA	A	C	C	G	NA	A	G
27	G	G	A	C	C	G	A	A	A
28	G	NA	A	C	C	A	A	A	A
29	G	G	A	C	C	A	A	A	G
31	T	NA	A	C	C	A	A	A	A
33	G	NA	NA	C	C	A	NA	A	A
34	G	NA	A	C	C	G	A	A	A
35	G	G	A	C	NA	NA	A	A	A
36	G	NA	NA	C	NA	G	A	A	NA
37	NA	NA	NA	C	C	G	A	A	G
38	G	NA	NA	C	C	G	A	A	G
39	T	NA	G	C	C	G	NA	A	NA
40	G	NA	NA	C	C	A	NA	A	NA
41	G	NA	NA	C	C	A	NA	NA	NA
42	G	NA	NA	C	C	A	A	NA	NA
43	G	NA	NA	C	C	G	A	NA	NA
44	G	G	NA	C	C	G	A	NA	NA
45	G	G	NA	C	NA	G	A	A	NA
46	G	G	NA	C	C	A	NA	NA	NA
47	G	NA	NA	C	NA	A	NA	NA	NA
48	NA	NA	NA	C	C	NA	NA	NA	NA
49	G	NA	NA	C	C	G	A	A	NA
50	G	NA	NA	C	C	G	A	NA	NA
51	G	NA	NA	C	C	NA	NA	NA	NA
52	T	NA	NA	C	C	A	A	NA	NA
53	G	G	NA	C	C	NA	A	A	NA
54	T	NA	NA	C	C	G	A	A	G
55	G	NA	NA	C	T	G	A	A	NA
56	G	NA	NA	C	C	G	A	A	NA
57	G	G	A	C	C	NA	A	A	A
58	T	A	A	C	C	G	NA	A	A
59	G	NA	NA	NA	C	NA	NA	A	G
60	G	A	G	C	C	A	NA	A	A
61	NA	NA	A	NA	NA	NA	NA	A	A
62	G	NA	G	C	C	G	A	A	G
63	G	NA	A	C	C	NA	A	A	G
64	G	NA	A	C	C	A	NA	A	G
65	G	NA	NA	C	C	A	NA	A	NA
68	T	A	A	C	C	G	A	A	G
69	G	NA	A	C	C	A	NA	A	A
70	G	G	A	C	C	A	NA	A	A
71	T	NA	G	C	NA	G	NA	A	A
73	G	G	A	C	C	G	A	A	G
74	G	G	G	C	NA	G	A	A	A
75	G	G	A	C	C	G	A	A	A
76	T	A	NA	C	C	G	A	NA	NA
79	G	NA	G	C	C	G	A	A	A
80	T	A	A	C	T	G	A	A	A
81	T	A	G	C	C	G	A	A	G
84	T	A	A	C	C	G	A	A	G
86	G	NA	A	C	C	G	A	A	A
87	NA	NA	A	C	C	NA	NA	A	A
88	NA	NA	NA	C	C	NA	NA	A	NA
89	G	G	A	C	C	A	C	A	A
90	G	G	G	T	C	A	C	A	G
91	G	NA	A	C	C	A	A	A	G
92	T	NA	A	C	C	G	NA	A	A
93	NA	NA	G	C	C	G	A	A	G
94	T	NA	G	C	NA	G	A	A	A
95	G	NA	G	T	C	A	NA	A	A
96	G	NA	NA	C	C	NA	A	A	A
97	G	NA	A	C	C	A	C	A	G

Table A.3 (continued)

Line \ SNP	Pvf1Exon1_C	Pvf1Exon1_C	Pvf1Exon4_C	Pvf1Exon4_C	Pvf1Exon5_C	Pvf1Intron1_	Pvf1Intron1_	Pvf1Intron4_	Rps65inter_C
98	T	A	G	C	C	G	A	A	G
99	T	A	A	C	C	G	C	A	G
101	G	NA	A	C	C	G	A	A	A
102	G	NA	A	C	NA	NA	A	A	A
103	G	NA	A	C	C	A	C	A	A
104	T	A	G	C	C	G	A	A	A
105	G	G	A	C	C	A	NA	A	A
106	T	NA	NA	C	C	G	NA	A	A
107	G	NA	A	C	C	A	C	A	A
108	G	A	NA	C	C	A	C	A	G
109	G	NA	A	C	NA	A	C	A	NA
110	NA	NA	NA	C	C	NA	C	A	A
111	G	G	NA	C	C	G	A	NA	NA
112	NA	NA	NA	C	C	G	A	NA	A
113	G	NA	A	C	C	G	A	A	G
114	G	G	G	C	C	G	A	A	G
115	T	NA	G	C	C	G	C	A	G
116	NA	NA	A	C	C	A	C	A	A
117	G	NA	A	C	C	A	C	A	G
119	NA	NA	G	C	C	G	A	A	A
122	G	A	A	C	C	A	C	A	A
123	T	A	A	C	T	A	A	G	G
125	T	A	G	C	C	G	C	A	G
126	G	NA	A	C	C	G	A	A	A
127	G	NA	G	C	C	G	C	A	G
128	G	G	A	C	T	G	A	A	A
130	G	NA	G	C	C	G	A	A	G
131	G	NA	A	C	C	A	NA	A	A
134	T	A	G	C	C	G	A	A	A
136	NA	NA	A	T	C	A	NA	NA	G
137	T	NA	A	NA	C	G	C	A	G
138	G	NA	A	C	C	A	C	A	A
139	NA	NA	NA	C	C	G	NA	A	G
140	T	NA	G	C	C	G	A	A	G
142	T	A	A	C	C	G	A	A	A
143	T	A	A	C	C	G	A	A	A
144	T	NA	G	C	C	G	A	A	G
145	G	NA	G	C	C	G	A	A	A
146	G	G	A	C	C	G	A	A	A
148	T	A	A	C	C	A	A	A	G
149	G	G	A	C	C	A	NA	A	A
151	G	A	A	C	C	G	C	A	A
152	T	NA	G	C	C	G	A	A	A
153	T	NA	A	C	T	A	A	A	G
154	G	NA	A	C	C	A	NA	A	A
155	G	NA	A	C	C	A	NA	A	A
158	G	NA	A	C	C	G	NA	A	G
160	G	NA	G	C	T	A	NA	A	G
164	G	NA	G	C	C	A	NA	A	G
166	G	NA	G	C	C	G	NA	A	G
167	G	G	A	T	NA	A	NA	G	A
168	G	G	A	C	C	G	A	A	A
169	G	NA	A	C	C	A	NA	A	G
172	G	NA	A	C	C	G	C	A	G
173	NA	G	NA	NA	NA	NA	NA	NA	G
174	T	NA	G	C	C	G	A	A	A
201	T	NA	G	C	C	G	A	A	G
202	T	NA	G	C	C	G	A	A	A
203	NA	NA	A	C	C	A	NA	A	G
204	NA	NA	A	C	C	G	A	A	NA
205	G	NA	A	C	C	A	A	A	G
206	T	A	G	C	C	G	NA	A	A
207	G	NA	G	C	C	G	A	A	A
208	G	G	A	C	C	G	A	A	A
209	T	NA	G	C	C	G	A	A	A
210	G	G	A	C	C	A	A	A	G
211	G	NA	G	C	C	G	A	A	G
212	NA	NA	NA	C	C	G	A	A	A
213	T	NA	G	C	C	A	A	A	G
214	G	NA	A	C	C	A	NA	A	A
215	G	A	G	C	C	A	NA	A	G
216	T	NA	A	C	T	G	NA	A	A
217	G	NA	A	C	T	G	A	A	A
218	G	NA	A	C	C	A	C	A	A
219	G	NA	A	C	C	A	NA	A	A
220	G	NA	G	C	T	A	C	A	NA
221	NA	NA	A	C	C	G	NA	A	NA
222	T	NA	A	C	C	G	A	A	A
223	G	NA	A	C	C	A	A	A	G
224	G	A	A	C	C	A	C	A	G
225	G	NA	G	C	C	G	A	A	G

Table A.3 (continued)

Line \ SNP	Rps6Exon3_(	Rps6Intron1	Ser75inter_(	Ser7Exon3_(	Ser7Exon3_(	Tak15UTR_0	Tak1Exon2_(	Tak1Exon3_(	Tak1Exon4_(
1	T	A	C	A	T	NA	A	A	C
2	G	A	C	NA	T	NA	NA	NA	NA
3	G	T	C	A	T	NA	T	G	NA
4	G	T	C	A	T	G	A	NA	NA
6	NA	NA	C	A	T	NA	NA	G	NA
7	G	NA	C	A	A	NA	NA	NA	NA
8	G	NA	C	A	T	NA	T	G	NA
9	G	A	C	A	A	G	A	A	C
10	G	A	C	A	T	NA	T	G	T
11	G	T	C	A	T	NA	A	A	T
12	G	A	C	A	A	NA	A	A	C
13	G	A	C	A	T	NA	A	A	C
14	G	A	C	A	T	G	A	A	C
15	G	T	C	A	T	T	T	G	T
16	G	T	C	C	A	T	T	G	T
17	G	A	C	A	T	G	A	A	C
22	G	A	C	A	T	NA	T	NA	C
23	G	T	C	C	T	G	T	A	C
24	G	A	C	A	T	T	T	NA	NA
25	G	A	C	A	NA	T	A	A	C
26	G	A	C	A	T	T	A	A	C
27	G	A	C	A	T	T	T	G	T
28	G	A	C	A	T	G	T	G	T
29	G	NA	C	C	T	G	A	A	C
31	G	A	C	A	T	G	A	G	C
33	G	A	C	A	T	T	A	A	C
34	G	A	C	A	T	G	A	A	C
35	G	A	C	A	T	T	A	NA	NA
36	G	NA	C	A	A	NA	A	NA	C
37	G	T	C	A	T	T	A	NA	NA
38	G	NA	C	A	T	G	A	NA	NA
39	G	A	C	A	T	T	A	G	T
40	G	T	C	A	T	G	A	A	C
41	G	NA	C	A	T	NA	A	A	NA
42	G	A	C	A	T	NA	T	G	T
43	G	NA	C	A	T	NA	A	A	C
44	G	NA	C	A	T	NA	A	A	C
45	G	A	C	A	T	NA	A	A	C
46	G	A	C	A	T	NA	A	A	C
47	G	NA	C	A	T	NA	T	G	T
48	G	A	C	A	A	NA	T	NA	C
49	G	NA	C	A	NA	G	A	A	C
50	G	A	C	A	T	NA	T	A	C
51	G	A	C	A	T	NA	A	A	C
52	G	T	C	A	T	NA	A	G	T
53	G	A	C	A	NA	NA	A	A	C
54	G	T	C	A	T	G	A	NA	T
55	G	A	C	A	A	G	A	G	T
56	G	A	C	A	T	NA	A	A	C
57	G	A	C	A	NA	NA	NA	G	T
58	G	A	C	A	T	T	T	G	NA
59	G	T	C	A	T	T	A	A	C
60	G	A	C	A	T	G	T	G	C
61	NA	NA	NA	NA	NA	T	NA	NA	NA
62	G	NA	C	A	T	G	A	A	C
63	G	NA	C	A	A	G	A	A	C
64	G	T	C	C	A	G	A	A	C
65	NA	NA	C	A	T	T	T	G	NA
68	G	A	C	A	T	G	A	A	C
69	G	A	C	C	A	T	A	A	C
70	G	A	C	A	A	T	T	NA	NA
71	G	A	C	A	T	G	A	A	C
73	G	NA	C	A	T	T	T	G	NA
74	G	A	C	A	T	G	A	G	NA
75	G	A	C	A	A	G	A	G	C
76	G	A	C	A	T	NA	A	A	C
79	G	A	C	A	A	G	A	A	C
80	G	A	C	A	A	G	A	G	C
81	G	NA	C	A	T	T	A	A	C
84	G	A	C	A	T	T	A	G	T
86	G	A	C	A	T	G	A	A	C
87	G	A	C	A	T	G	A	A	C
88	G	NA	C	A	NA	NA	T	G	T
89	G	A	C	NA	NA	G	A	A	C
90	G	NA	C	C	A	T	T	G	T
91	G	T	C	A	A	T	A	G	NA
92	T	A	C	A	T	G	A	A	C
93	T	NA	C	A	T	NA	A	NA	NA
94	G	A	C	A	A	G	A	A	C
95	G	A	C	C	T	G	A	G	T
96	G	A	C	A	T	G	A	NA	NA
97	G	NA	C	C	T	G	A	G	C

Table A.3 (continued)

Line \ SNP	Rps6Exon3_(	Rps6Intron1	Ser75inter_(	Ser7Exon3_(	Ser7Exon3_(	Tak15UTR_0	Tak1Exon2_(	Tak1Exon3_(	Tak1Exon4_(
98	G	T	C	A	T	G	T	A	T
99	G	T	C	A	T	G	A	A	C
101	G	A	C	A	T	G	T	G	C
102	G	A	C	C	A	G	A	G	T
103	G	A	C	C	T	G	A	A	C
104	G	A	C	A	T	T	A	G	T
105	G	A	C	A	T	T	T	G	NA
106	NA	A	C	A	NA	G	A	G	C
107	G	A	C	A	T	G	A	A	C
108	G	T	C	A	T	G	T	G	C
109	G	NA	C	A	T	T	NA	NA	T
110	G	A	C	A	T	T	A	A	C
111	G	A	C	A	T	NA	A	A	C
112	G	A	C	A	T	T	A	A	NA
113	G	A	C	A	A	G	T	G	T
114	G	T	C	A	A	G	T	G	T
115	G	T	C	A	A	T	T	G	T
116	G	A	NA	A	T	G	NA	G	T
117	G	A	C	A	T	T	T	G	NA
119	G	A	C	A	NA	T	T	G	C
122	G	A	C	C	A	G	A	A	C
123	G	T	C	C	T	T	T	G	C
125	G	T	C	A	A	G	A	G	NA
126	G	A	G	A	A	T	T	G	T
127	G	A	C	A	A	G	A	A	C
128	G	A	C	C	T	T	A	A	C
130	G	A	C	A	NA	G	A	G	T
131	G	A	C	A	T	G	A	A	C
134	G	A	C	A	T	T	A	A	C
136	G	A	C	A	T	NA	A	A	C
137	G	T	NA	A	T	G	NA	G	T
138	G	A	C	C	A	G	A	A	C
139	G	A	C	A	A	G	A	A	C
140	G	T	C	A	A	T	A	A	C
142	NA	A	C	NA	NA	T	A	A	C
143	G	A	C	A	A	G	T	G	C
144	G	T	C	A	T	G	A	A	C
145	G	A	C	A	T	T	T	G	T
146	T	A	C	A	T	G	A	A	C
148	G	T	C	A	T	G	A	G	C
149	G	A	C	A	T	G	A	A	C
151	G	A	C	A	A	G	A	G	NA
152	G	A	C	A	T	G	T	G	T
153	G	T	C	A	A	G	A	A	C
154	G	A	C	A	A	G	A	G	T
155	G	A	C	A	A	T	A	A	C
158	G	A	C	A	A	T	A	A	NA
160	G	T	C	A	T	G	A	A	C
164	G	T	C	A	A	G	A	G	T
166	G	T	C	C	T	G	A	A	C
167	G	A	C	A	T	T	A	G	T
168	G	A	C	A	T	T	T	G	NA
169	G	A	C	A	A	G	A	G	NA
172	G	T	G	A	A	G	A	A	C
173	G	A	C	A	A	G	NA	NA	C
174	G	A	C	A	A	G	A	A	C
201	G	T	C	A	A	T	A	A	C
202	G	A	C	A	T	G	A	G	NA
203	G	T	C	A	T	G	A	A	C
204	G	NA	C	NA	T	G	NA	G	NA
205	G	T	C	C	T	G	A	G	T
206	G	A	C	A	T	G	A	A	C
207	T	A	C	A	T	G	A	A	C
208	G	A	C	A	T	G	NA	NA	NA
209	G	A	C	A	T	G	T	G	NA
210	G	T	C	A	T	G	A	A	C
211	G	A	G	A	T	T	T	G	T
212	G	A	NA	A	NA	G	A	A	C
213	G	T	C	C	T	G	A	A	C
214	G	A	C	A	T	G	A	A	C
215	G	T	C	A	A	G	A	A	C
216	G	A	C	A	T	G	A	A	C
217	G	A	C	A	T	G	A	A	C
218	G	A	C	C	A	G	A	A	C
219	G	A	C	A	T	G	A	G	C
220	G	NA	C	NA	T	G	A	G	NA
221	G	NA	C	NA	T	T	T	G	T
222	G	A	C	A	T	G	A	A	C
223	G	T	C	A	NA	G	T	G	T
224	G	A	C	A	A	G	A	G	NA
225	G	A	C	A	A	G	A	A	C

Table A.3 (continued)

Line \ SNP	Tak1Intron1	Traf25inter_	Traf2Exon1_	Traf2Exon1_	Traf3Exon2_	Traf3Exon3_	Tsf15inter_0	Tsf1Exon2_0	Tsf1Exon3_0
1	NA	NA	NA	NA	C	NA	NA	NA	NA
2	NA	NA	NA	NA	C	NA	NA	NA	T
3	NA	NA	NA	C	C	NA	T	NA	NA
4	NA	NA	NA	C	C	NA	C	C	T
6	G	G	NA	C	C	A	C	NA	T
7	G	G	G	C	C	G	NA	C	T
8	NA	NA	NA	NA	C	NA	NA	NA	T
9	NA	NA	G	C	C	A	C	NA	T
10	NA	NA	NA	NA	G	NA	T	NA	NA
11	NA	NA	NA	C	C	NA	C	NA	NA
12	NA	NA	NA	C	C	NA	C	NA	NA
13	NA	NA	NA	C	C	NA	C	NA	NA
14	G	NA	G	C	C	NA	NA	C	NA
15	G	G	A	C	C	G	C	C	T
16	G	A	G	T	C	G	C	C	T
17	G	NA	G	C	C	G	C	T	T
22	NA	NA	NA	C	C	NA	C	NA	NA
23	T	G	G	C	C	A	T	C	T
24	NA	NA	G	C	C	A	T	C	T
25	G	NA	G	NA	C	A	C	C	T
26	G	G	G	C	C	NA	C	C	T
27	G	G	G	C	C	A	C	C	T
28	G	G	G	C	C	G	T	C	T
29	G	G	G	C	C	G	C	C	T
31	G	G	A	C	C	A	T	C	T
33	G	G	A	C	C	A	C	C	T
34	G	NA	G	C	C	NA	T	T	T
35	G	G	G	C	C	NA	C	C	T
36	NA	NA	G	C	C	NA	C	NA	T
37	G	G	G	C	C	A	C	C	T
38	G	G	G	C	C	A	C	C	NA
39	G	NA	G	C	C	A	C	C	T
40	NA	NA	G	C	C	A	C	C	T
41	NA	NA	NA	C	C	NA	C	NA	NA
42	NA	NA	NA	C	C	NA	T	NA	NA
43	NA	NA	NA	NA	C	NA	T	NA	NA
44	NA	NA	NA	NA	C	NA	NA	NA	NA
45	NA	NA	G	C	C	NA	C	NA	T
46	NA	NA	NA	C	C	NA	C	NA	NA
47	NA	NA	NA	C	C	NA	C	NA	NA
48	NA	NA	NA	C	C	NA	C	NA	NA
49	G	A	G	T	C	NA	C	C	C
50	NA	NA	NA	C	C	NA	C	NA	NA
51	NA	NA	NA	NA	C	NA	NA	NA	NA
52	NA	NA	NA	C	C	NA	C	NA	NA
53	NA	NA	G	C	C	NA	T	NA	T
54	G	A	G	T	C	A	C	C	T
55	NA	NA	G	C	C	NA	C	NA	T
56	NA	NA	NA	C	C	NA	T	NA	NA
57	G	NA	G	C	C	A	NA	T	NA
58	G	G	G	C	C	G	C	C	T
59	G	G	G	C	C	NA	C	C	T
60	G	G	G	C	C	G	C	C	T
61	G	G	G	NA	NA	A	NA	C	T
62	G	A	G	T	C	G	C	C	C
63	G	A	G	T	C	NA	C	C	T
64	G	G	G	C	C	G	T	C	T
65	G	G	G	C	C	A	C	C	NA
68	G	G	G	C	C	G	T	C	T
69	G	G	G	C	C	A	T	C	T
70	G	G	G	C	C	A	T	C	T
71	G	G	G	C	C	A	T	C	T
73	G	A	G	T	C	A	T	C	T
74	G	G	G	C	C	A	C	C	T
75	G	G	G	C	C	G	C	C	T
76	NA	NA	NA	C	C	NA	C	NA	NA
79	G	G	G	C	C	G	T	C	T
80	G	G	G	C	C	G	T	C	T
81	G	G	G	C	C	NA	C	C	T
84	G	NA	G	C	C	A	C	C	T
86	G	A	G	T	C	G	T	C	T
87	G	NA	G	T	C	A	NA	C	T
88	NA	NA	NA	C	C	NA	NA	NA	NA
89	G	NA	G	C	C	A	C	T	C
90	G	G	G	C	C	A	C	C	T
91	G	G	G	C	C	G	C	C	T
92	G	G	G	C	C	G	C	C	T
93	G	NA	G	C	C	A	NA	C	T
94	G	G	G	C	C	A	C	T	T
95	G	NA	G	C	C	G	T	C	T
96	G	NA	G	C	C	G	NA	C	NA
97	G	NA	G	C	C	A	T	C	T

Table A.3 (continued)

Line \ SNP	Tak1Intron1	Traf25inter_	Traf2Exon1_	Traf2Exon1_	Traf3Exon2_	Traf3Exon3_	Tsf15inter_0	Tsf1Exon2_0	Tsf1Exon3_0
98	G	A	G	T	C	A	C	C	T
99	G	A	G	T	C	A	C	C	T
101	G	A	G	T	C	A	C	C	T
102	G	G	G	C	C	A	C	C	T
103	G	G	G	C	C	A	C	C	T
104	G	G	G	C	C	A	C	C	T
105	G	G	G	C	C	G	T	C	T
106	NA	G	NA	C	C	A	T	NA	T
107	G	NA	G	C	C	A	NA	C	T
108	G	NA	G	T	C	A	C	C	T
109	G	NA	G	C	C	A	NA	NA	T
110	G	A	G	T	C	G	NA	C	T
111	NA	NA	NA	C	C	NA	T	NA	NA
112	NA	NA	NA	T	C	NA	T	NA	NA
113	G	G	G	C	C	A	C	T	C
114	G	G	G	C	C	G	T	NA	T
115	G	G	G	C	C	A	C	C	T
116	G	G	G	C	C	A	NA	C	C
117	G	G	G	C	C	G	T	C	T
119	G	G	G	NA	C	G	C	C	C
122	G	G	G	C	C	A	T	C	T
123	G	G	G	C	C	A	NA	C	T
125	G	A	G	T	C	A	T	C	T
126	G	G	G	C	C	A	C	C	T
127	G	G	G	C	C	A	T	C	T
128	G	A	G	T	C	A	T	C	T
130	G	G	G	C	C	A	C	C	T
131	G	G	G	C	C	G	T	C	T
134	G	G	G	C	C	A	NA	C	T
136	G	G	G	C	C	A	NA	C	T
137	G	G	G	NA	NA	A	C	C	T
138	G	G	G	C	C	A	T	C	T
139	G	A	G	T	C	G	C	C	NA
140	G	A	G	T	C	A	C	T	T
142	G	G	G	C	C	A	T	C	T
143	G	G	G	C	C	A	C	C	T
144	G	G	G	C	C	G	C	C	NA
145	G	G	G	C	C	A	NA	C	T
146	G	G	G	C	C	G	C	C	T
148	G	G	G	C	C	A	T	C	T
149	G	A	G	T	C	G	C	T	C
151	G	G	G	C	C	A	C	C	C
152	G	A	G	T	C	A	T	C	T
153	G	G	G	C	C	A	C	C	T
154	G	G	G	C	C	A	C	C	T
155	G	NA	G	T	C	A	C	C	T
158	G	G	G	NA	C	G	C	C	T
160	G	A	G	T	C	A	NA	C	T
164	G	G	A	C	C	A	T	C	T
166	G	G	G	C	C	G	T	C	T
167	G	G	G	C	C	A	T	C	T
168	G	G	G	C	C	A	T	C	T
169	G	G	G	C	C	G	NA	C	T
172	G	G	G	C	C	A	C	C	C
173	G	NA	G	NA	C	NA	NA	C	T
174	G	G	G	C	C	A	C	C	T
201	G	G	G	C	C	A	C	C	T
202	G	G	G	C	C	A	T	C	T
203	G	G	G	C	C	A	T	C	T
204	G	G	G	C	C	A	NA	T	C
205	G	G	G	C	C	A	C	T	C
206	G	G	G	C	C	A	C	C	T
207	G	G	G	C	C	A	C	C	T
208	G	G	G	C	C	A	C	C	T
209	G	G	G	C	C	A	NA	C	T
210	G	NA	G	T	C	A	T	C	T
211	G	G	G	C	G	G	T	C	C
212	G	G	G	C	C	NA	NA	C	T
213	G	G	G	C	C	A	NA	C	T
214	G	G	G	C	C	G	C	C	T
215	G	G	G	C	C	A	C	C	T
216	G	G	G	C	C	G	C	C	T
217	G	G	G	C	C	A	C	C	T
218	G	G	G	C	C	A	C	C	T
219	G	NA	G	T	C	G	NA	C	T
220	G	G	G	C	C	A	C	C	T
221	G	G	G	C	C	A	C	NA	C
222	G	G	G	C	C	A	T	C	T
223	G	NA	G	T	C	A	NA	C	T
224	G	G	G	C	C	A	C	C	T
225	G	G	G	C	C	NA	C	C	T

Table A.3 (continued)

Line \ SNP	Tsf1Exon3_0	Tsf1Exon3_0	Tsf1Exon3_0	upd2Exon1_	upd2Exon3_	upd2Intron1	upd3Exon1_(	upd3Exon2_(	upd3Exon3_	upd3Intron1
1	NA	NA	NA	NA	NA	NA	G	NA	G	G
2	T	C	NA	NA	NA	A	G	NA	G	NA
3	NA	NA	T	G	T	NA	G	NA	G	G
4	T	C	NA	NA	G	G	G	T	G	G
6	T	C	NA	NA	NA	NA	G	T	G	G
7	NA	C	NA	NA	NA	NA	G	T	G	G
8	T	C	NA	NA	NA	NA	G	NA	A	G
9	T	C	G	G	T	A	G	C	G	G
10	NA	NA	NA	A	G	NA	A	NA	A	T
11	NA	NA	G	G	T	NA	G	NA	G	G
12	NA	NA	T	G	G	NA	G	NA	G	G
13	NA	NA	NA	G	T	NA	G	NA	G	G
14	C	C	T	A	G	G	G	T	G	G
15	T	C	G	G	T	G	G	T	G	G
16	T	C	T	A	G	G	G	T	G	G
17	T	C	G	G	G	G	G	T	G	G
22	NA	NA	G	G	G	NA	G	NA	G	G
23	C	C	G	A	G	G	G	C	A	G
24	C	C	G	G	T	A	G	T	G	G
25	C	C	NA	G	T	G	G	T	G	G
26	T	C	T	G	G	G	G	T	G	G
27	T	C	G	G	T	A	G	T	G	T
28	C	C	T	G	NA	A	G	T	G	G
29	T	C	T	G	T	A	G	T	G	G
31	C	C	G	A	G	G	G	T	G	G
33	C	C	T	G	T	A	G	T	G	G
34	C	C	T	G	NA	G	G	T	G	G
35	T	C	NA	NA	G	G	G	T	G	G
36	T	C	G	NA	G	NA	G	NA	G	G
37	C	C	T	NA	NA	NA	G	T	G	G
38	NA	C	T	G	T	A	G	NA	G	G
39	NA	C	NA	G	T	A	G	NA	G	NA
40	T	C	G	G	NA	A	G	T	G	G
41	NA	NA	T	G	T	NA	G	NA	G	G
42	NA	NA	T	G	NA	NA	G	NA	G	G
43	NA	C	G	A	G	NA	G	NA	G	G
44	NA	NA	G	NA	NA	NA	G	NA	G	G
45	T	C	G	G	T	A	G	T	G	G
46	NA	NA	G	G	T	NA	G	NA	G	G
47	NA	C	T	G	T	NA	G	NA	G	G
48	NA	NA	T	NA	G	NA	G	NA	G	G
49	C	C	T	G	NA	A	G	T	G	G
50	NA	NA	T	G	T	NA	G	NA	G	G
51	NA	NA	T	A	G	NA	G	NA	G	G
52	NA	NA	G	G	T	NA	G	NA	G	T
53	NA	C	G	A	G	NA	G	NA	G	G
54	T	C	T	A	G	G	G	C	G	G
55	T	C	G	G	G	NA	G	T	G	G
56	NA	C	T	G	T	NA	G	NA	G	G
57	C	C	NA	NA	T	A	G	NA	G	G
58	T	C	T	A	G	G	G	T	G	G
59	T	C	NA	NA	G	G	G	NA	G	G
60	T	C	T	G	T	A	G	T	G	G
61	T	C	NA	NA	NA	G	NA	NA	NA	NA
62	C	C	T	G	T	A	G	T	G	G
63	T	C	NA	NA	NA	NA	G	T	G	G
64	C	C	T	A	G	G	G	T	G	G
65	NA	C	T	NA	NA	NA	G	NA	G	G
68	C	C	G	G	NA	A	G	T	G	G
69	T	C	T	G	T	A	G	T	G	G
70	C	C	G	G	T	A	G	T	G	G
71	C	C	NA	G	T	A	G	T	G	G
73	C	C	T	G	T	A	G	T	G	G
74	T	C	T	G	T	A	G	T	G	G
75	T	C	T	G	T	A	G	T	G	G
76	NA	NA	T	A	G	NA	G	NA	G	G
79	C	C	G	NA	G	G	G	T	G	G
80	C	C	T	A	G	G	A	C	A	G
81	C	C	T	G	G	G	G	T	G	G
84	T	C	G	G	G	G	G	T	G	G
86	C	C	G	G	T	A	G	T	G	G
87	T	C	NA	G	T	NA	G	T	G	G
88	NA	NA	G	NA	NA	NA	G	NA	G	G
89	T	C	G	G	G	G	G	T	G	G
90	T	C	T	G	G	G	G	C	G	G
91	T	C	G	A	G	G	G	C	G	G
92	T	C	T	A	G	G	G	T	G	G
93	T	C	NA	NA	NA	NA	G	T	G	G
94	T	C	G	A	G	G	G	C	G	G
95	C	C	G	A	G	G	G	T	G	G
96	NA	NA	NA	NA	NA	NA	NA	T	G	G
97	C	C	NA	G	T	A	G	T	A	G

Table A.3 (continued)

Line \ SNP	Tsf1Exon3_0	Tsf1Exon3_0	Tsf1Exon3_0	upd2Exon1_	upd2Exon3_	upd2Intron1	upd3Exon1_(	upd3Exon2_(	upd3Exon3_	upd3Intron1
98	T	C	T	A	G	G	G	C	G	G
99	T	C	T	A	G	G	G	T	G	G
101	T	C	G	A	G	G	G	T	G	G
102	T	C	G	G	G	G	G	T	G	G
103	T	C	G	NA	NA	A	G	T	G	G
104	C	C	T	A	G	G	G	C	G	G
105	T	C	G	G	NA	A	A	T	G	G
106	C	C	T	G	T	A	NA	NA	NA	NA
107	C	C	NA	G	NA	A	G	T	G	G
108	T	C	G	A	G	G	G	T	G	G
109	NA	C	G	G	T	NA	G	NA	G	G
110	NA	NA	G	G	T	A	G	T	G	G
111	NA	NA	G	A	G	NA	G	T	G	G
112	NA	NA	G	G	T	NA	G	T	G	G
113	C	NA	T	A	T	A	G	C	G	G
114	C	NA	G	A	G	G	G	T	G	G
115	T	NA	G	G	T	A	G	T	G	G
116	C	NA	T	A	G	G	G	T	G	G
117	C	NA	G	G	T	A	G	T	G	G
119	C	NA	T	A	G	G	G	C	G	G
122	C	NA	T	A	G	G	G	T	G	G
123	T	NA	G	G	T	G	G	T	G	G
125	C	NA	T	A	G	G	G	T	G	G
126	C	NA	T	G	T	A	G	T	G	G
127	C	NA	G	G	T	A	G	C	G	G
128	C	NA	T	G	T	G	G	T	G	G
130	T	NA	T	G	T	G	G	T	G	G
131	C	NA	G	A	G	G	G	T	G	T
134	C	NA	T	G	G	G	G	T	G	G
136	NA	NA	NA	G	T	A	G	T	G	NA
137	T	NA	G	A	G	G	G	C	A	T
138	T	NA	T	G	T	A	G	T	G	G
139	T	NA	T	NA	NA	NA	G	T	G	G
140	T	NA	G	A	G	G	G	T	G	G
142	C	NA	G	G	T	A	G	T	G	G
143	T	NA	T	G	G	G	G	T	G	G
144	T	NA	G	G	T	G	G	T	G	G
145	C	NA	G	G	T	A	G	C	G	T
146	T	NA	G	G	G	G	G	T	G	G
148	C	NA	G	A	G	G	G	T	G	G
149	T	NA	G	G	G	G	G	T	G	G
151	C	NA	T	G	T	A	G	T	G	G
152	C	NA	T	NA	G	G	G	T	G	G
153	C	NA	T	G	T	A	G	T	G	T
154	T	NA	G	G	T	A	G	T	G	G
155	T	NA	T	A	G	G	G	T	G	G
158	T	NA	T	G	T	A	G	C	G	G
160	T	NA	G	G	T	A	G	T	G	G
164	C	NA	T	A	G	G	G	C	G	G
166	C	NA	G	G	T	A	G	T	G	G
167	T	NA	G	G	T	A	G	C	G	T
168	C	NA	G	G	T	A	G	C	G	G
169	C	NA	T	A	G	G	G	C	G	G
172	C	NA	T	A	G	G	G	C	G	G
173	NA	NA	NA	NA	NA	NA	G	NA	G	NA
174	T	NA	G	G	T	A	G	T	G	G
201	T	NA	G	G	G	G	G	T	G	G
202	C	NA	T	G	G	G	G	T	G	G
203	C	NA	G	G	T	A	G	C	G	T
204	C	NA	G	A	NA	G	G	C	G	G
205	C	NA	G	G	T	A	G	T	G	G
206	T	NA	G	G	T	A	G	T	G	G
207	C	NA	G	G	T	A	G	C	G	T
208	T	NA	G	A	G	G	G	T	G	G
209	T	NA	T	G	T	A	G	T	G	G
210	C	NA	T	G	T	A	G	C	G	T
211	C	NA	T	G	T	A	G	T	G	T
212	NA	NA	NA	G	NA	NA	G	T	G	G
213	T	NA	G	G	T	A	G	T	G	G
214	T	NA	G	G	T	A	G	T	G	G
215	T	NA	G	G	T	A	G	T	G	T
216	T	NA	T	G	T	A	G	T	G	G
217	T	NA	G	G	G	G	G	T	G	G
218	T	NA	G	G	T	A	G	C	G	T
219	C	NA	G	G	T	A	G	C	G	G
220	T	NA	G	G	T	A	G	T	G	G
221	NA	NA	NA	NA	NA	NA	G	T	G	G
222	T	NA	G	G	T	A	G	C	G	T
223	T	NA	G	A	G	G	G	T	G	G
224	T	NA	G	G	T	A	G	T	G	G
225	T	NA	T	G	T	A	G	C	G	G

Table A.4 TaqMan Primer and Probe Sequences

Gene		Sequence
Def	MGB probe	AGGATGCCCACCAGGA
	forward primer	GAGGATCATGTCCTGGTGCAT
	reverse primer	TCGCTTCTGGCGGCTATG
DptA	MGB probe	TTTGCAGTCCAGGGTC
	forward primer	GCGGCGATGGTTTTGG
	reverse primer	CGCTGGTCCACACCTTCTG
Mtk	MGB probe	GCTGGGTGTGATGG
	forward primer	AACTTAATCTTGGAGCGATTTTTCTG
	reverse primer	ACGGCCTCGTATCGAAAATG
PGRP-SA	MGB probe	CGAAGGCACTGGTTG
	forward primer	TCGGCAACGATGGTATCGTA
	reverse primer	GGCACCGCGCAATCC
Tsf1	MGB probe	AGTGCCGCCTTCC
	forward primer	GAACGCAGCAGGACAAGGA
	reverse primer	CTGCTGCAGGGTGCGAAT
RpL32	MGB probe	AGCTGTCGCACAAAT
	forward primer	AGGCCCAAGATCGTGAAGAA
	reverse primer	GACGCACTCTGTTGTCGATACG

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Table B.1 Genes Included on BeadChips

Category	FBgn	Symbol	Function	Category	FBgn	Symbol	Function
immunity	FBgn0012042	AttA	effector	other	FBgn0058370	CG40370	\N
immunity	FBgn0041581	AttB	effector	other	FBgn0025628	CG4199	\N
immunity	FBgn0041579	AttC	effector	other	FBgn0036639	CG4229	\N
immunity	FBgn0038530	AttD	effector	other	FBgn0034761	CG4250	\N
immunity	FBgn0000165	Bc	effector	other	FBgn0031405	CG4267	\N
immunity	FBgn0002022	Catsup	effector	other	FBgn0034741	CG4269	\N
immunity	FBgn0000276	CecA1	effector	other	FBgn0035589	CG4618	\N
immunity	FBgn0000277	CecA2	effector	other	FBgn0027584	CG4757	\N
immunity	FBgn0000278	CecB	effector	other	FBgn0028514	CG4793	\N
immunity	FBgn0000279	CecC	effector	other	FBgn0039434	CG5468	\N
immunity	FBgn0034539	CG11159	effector	other	FBgn0034364	CG5493	\N
immunity	FBgn0028526	CG15293	effector	other	FBgn0034160	CG5550	\N
immunity	FBgn0032773	CG15825	effector	other	FBgn0034290	CG5773	\N
immunity	FBgn0035743	CG15829	effector	other	FBgn0038930	CG5778	\N
immunity	FBgn0029765	CG16756	effector	other	FBgn0040582	CG5791	\N
immunity	FBgn0034538	CG16799	effector	other	FBgn0038682	CG5835	\N
immunity	FBgn0034162	CG6426	effector	other	FBgn0032973	CG6675	\N
immunity	FBgn0034092	CG7798	effector	other	FBgn0030086	CG7033	\N
immunity	FBgn0035813	CG8492	effector	other	FBgn0031941	CG7211	\N
immunity	FBgn0038465	CG8913	effector	other	FBgn0031940	CG7214	\N
immunity	FBgn0000422	Ddc	effector	other	FBgn0031723	CG7251	\N
immunity	FBgn0010385	Def	effector	other	FBgn0038631	CG7695	\N
immunity	FBgn0000486	Dox-A2	effector	other	FBgn0033596	CG7738	\N
immunity	FBgn0000487	Dox-A3	effector	other	FBgn0032025	CG7778	\N
immunity	FBgn0004240	Dpt	effector	other	FBgn0038576	CG7940	\N
immunity	FBgn0034407	DptB	effector	other	FBgn0033234	CG8791	\N
immunity	FBgn0010388	Dro	effector	other	FBgn0035208	CG9184	\N
immunity	FBgn0052279	dro2	effector	other	FBgn0032069	CG9468	\N
immunity	FBgn0052283	dro3	effector	other	FBgn0036659	CG9701	\N
immunity	FBgn0052282	dro4	effector	other	FBgn0032472	CG9928	\N
immunity	FBgn0035434	dro5	effector	other	FBgn0039593	CG9989	\N
immunity	FBgn0052268	dro6	effector	other	FBgn0035636	Cralbp	\N
immunity	FBgn0010381	Drs	effector	other	FBgn0010383	Cyp18a1	\N
immunity	FBgn0052274	Drs-l	effector	other	FBgn0015032	Cyp4c3	\N
immunity	FBgn0029167	Hml	effector	other	FBgn0035964	Dhpr	\N
immunity	FBgn0034329	IM1	effector	other	FBgn0037977	Ect3	\N
immunity	FBgn0033835	IM10	effector	other	FBgn0005660	Ets21C	\N
immunity	FBgn0025583	IM2	effector	other	FBgn0000715	Fmrf	\N
immunity	FBgn0034328	IM23	effector	other	FBgn0000721	for	\N
immunity	FBgn0040736	IM3	effector	other	FBgn0037724	Fst	\N
immunity	FBgn0040653	IM4	effector	other	FBgn0001112	Gld	\N
immunity	FBgn0004425	LysB	effector	other	FBgn0035167	Gr61a	\N
immunity	FBgn0004426	LysC	effector	other	FBgn0001230	Hsp68	\N
immunity	FBgn0004427	LysD	effector	other	FBgn0020416	Idgf1	\N
immunity	FBgn0004428	LysE	effector	other	FBgn0020415	Idgf2	\N
immunity	FBgn0004429	LysP	effector	other	FBgn0020414	Idgf3	\N
immunity	FBgn0004430	LysS	effector	other	FBgn0026415	Idgf4	\N
immunity	FBgn0004431	LysX	effector	other	FBgn0001281	janB	\N
immunity	FBgn0014865	Mtk	effector	other	FBgn0003358	Jon99Ci	\N
immunity	FBgn0005626	ple	effector	other	FBgn0020638	Lcp65Ag1	\N
immunity	FBgn0003162	Pu	effector	other	FBgn0020637	Lcp65Ag2	\N
immunity	FBgn0011722	Tig	effector	other	FBgn0040104	lectin-24A	\N
immunity	FBgn0028396	TotA	effector	other	FBgn0016675	Lectin-galC1	\N
immunity	FBgn0038838	TotB	effector	other	FBgn0020278	loco	\N
immunity	FBgn0044812	TotC	effector	other	FBgn0051092	LpR2	\N
immunity	FBgn0044811	TotF	effector	other	FBgn0002565	Lsp2	\N
immunity	FBgn0031701	TotM	effector	other	FBgn0013987	MAPk-Ak2	\N
immunity	FBgn0044810	TotX	effector	other	FBgn0002869	MtnB	\N
immunity	FBgn0044809	TotZ	effector	other	FBgn0011672	Mvl	\N
immunity	FBgn0022355	Tsf1	effector	other	FBgn0013305	Nmda1	\N
immunity	FBgn0036299	Tsf2	effector	other	FBgn0034509	Obp57c	\N
immunity	FBgn0034094	Tsf3	effector	other	FBgn0046878	Obp83cd	\N
immunity	FBgn0041710	yellow-f	effector	other	FBgn0037589	Obp85a	\N
immunity	FBgn0038105	yellow-f2	effector	other	FBgn0013307	Odc1	\N

Table B.1 (continued)

Category	FBgn	Symbol	Function	Category	FBgn	Symbol	Function
immunity	FBgn0033301	CG12780	recognition	other	FBgn0002997	ome	\N
immunity	FBgn0034511	CG13422	recognition	other	FBgn0037590	Or85b	\N
immunity	FBgn0051217	CG31217	recognition	other	FBgn0026192	par-6	\N
immunity	FBgn0031547	CG3212	recognition	other	FBgn0035089	Phk-3	\N
immunity	FBgn0015924	crq	recognition	other	FBgn0003328	scb	\N
immunity	FBgn0010435	emp	recognition	other	FBgn0015808	ScpX	\N
immunity	FBgn0040323	GNBP1	recognition	other	FBgn0051069	spn-D	\N
immunity	FBgn0040322	GNBP2	recognition	other	FBgn0003511	Sry-beta	\N
immunity	FBgn0040321	GNBP3	recognition	other	FBgn0031506	Tdp1	\N
immunity	FBgn0028430	He	recognition	other	FBgn0011584	Trp1	\N
immunity	FBgn0035975	PGRP-LA	recognition	other	FBgn0017456	Ubc84D	\N
immunity	FBgn0037906	PGRP-LB	recognition	other	FBgn0003961	Uro	\N
immunity	FBgn0035976	PGRP-LC	recognition	other	FBgn0011737	wee	\N
immunity	FBgn0035635	PGRP-LD	recognition	other	FBgn0015565	yin	\N
immunity	FBgn0030695	PGRP-LE	recognition	other	FBgn0010501	Dcp-1	apoptosis
immunity	FBgn0035977	PGRP-LF	recognition	other	FBgn0028381	decay	apoptosis
immunity	FBgn0030310	PGRP-SA	recognition	other	FBgn0000477	DNaseII	apoptosis
immunity	FBgn0043578	PGRP-SB1	recognition	other	FBgn0015946	grim	apoptosis
immunity	FBgn0043577	PGRP-SB2	recognition	other	FBgn0035049	Mmp1	apoptosis
immunity	FBgn0043576	PGRP-SC1a	recognition	other	FBgn0027609	morgue	apoptosis
immunity	FBgn0033327	PGRP-SC1b	recognition	other	FBgn0011706	rpr	apoptosis
immunity	FBgn0043575	PGRP-SC2	recognition	other	FBgn0023076	Clk	circadian
immunity	FBgn0035806	PGRP-SD	recognition	other	FBgn0025680	cry	circadian
immunity	FBgn0014033	Sr-CI	recognition	other	FBgn0023094	cyc	circadian
immunity	FBgn0020377	Sr-CII	recognition	other	FBgn0023178	Pdf	circadian
immunity	FBgn0020376	Sr-CIII	recognition	other	FBgn0016694	Pdp1	circadian
immunity	FBgn0041183	TepI	recognition	other	FBgn0003068	per	circadian
immunity	FBgn0041182	TepII	recognition	other	FBgn0014396	tim	circadian
immunity	FBgn0041181	TepIII	recognition	other	FBgn0039298	to	circadian
immunity	FBgn0041180	TepIV	recognition	other	FBgn0016076	vri	circadian
immunity	FBgn0004364	18w	signaling	other	FBgn0002626	RpL32	control
immunity	FBgn0000097	aop	signaling	other	FBgn0028683	spt4	control
immunity	FBgn0022131	aPKC	signaling	other	FBgn0000094	Anp	effector
immunity	FBgn0000173	ben	signaling	other	FBgn0039628	CG11841	effector
immunity	FBgn0038928	BG4	signaling	other	FBgn0031560	CG16713	effector
immunity	FBgn0000212	brm	signaling	other	FBgn0037678	CG16749	effector
immunity	FBgn0000229	bsk	signaling	other	FBgn0036024	CG18180	effector
immunity	FBgn0000250	cact	signaling	other	FBgn0032639	CG18563	effector
immunity	FBgn0032832	CG10662	signaling	other	FBgn0050088	CG30088	effector
immunity	FBgn0027930	CG1102	signaling	other	FBgn0031562	CG3604	effector
immunity	FBgn0039666	CG11501	signaling	other	FBgn0030449	CG4349	effector
immunity	FBgn0039102	CG16705	signaling	other	FBgn0039495	CG5909	effector
immunity	FBgn0037515	CG3066	signaling	other	FBgn0036969	CG6663	effector
immunity	FBgn0030925	CG6361	signaling	other	FBgn0027563	CG9631	effector
immunity	FBgn0030774	CG9675	signaling	other	FBgn0038211	CG9649	effector
immunity	FBgn0014141	cher	signaling	other	FBgn0015222	Fer1HCH	effector
immunity	FBgn0011274	Dif	signaling	other	FBgn0015221	Fer2LCH	effector
immunity	FBgn0000462	dl	signaling	other	FBgn0030092	fh	effector
immunity	FBgn0034738	Dnr1	signaling	other	FBgn0020906	Jon25Bi	effector
immunity	FBgn0020306	dom	signaling	other	FBgn0031654	Jon25Bii	effector
immunity	FBgn0043903	dome	signaling	other	FBgn0031653	Jon25Biii	effector
immunity	FBgn0020381	Dredd	signaling	other	FBgn0001285	Jon44E	effector
immunity	FBgn0010269	Dsor1	signaling	other	FBgn0035667	Jon65Ai	effector
immunity	FBgn0000533	ea	signaling	other	FBgn0035664	Jon65Aiv	effector
immunity	FBgn0028436	ECSIT	signaling	other	FBgn0039778	Jon99Fi	effector
immunity	FBgn0003731	Egfr	signaling	other	FBgn0019929	Ser7	effector
immunity	FBgn0020497	emb	signaling	other	FBgn0019928	Ser8	effector
immunity	FBgn0014179	gcm	signaling	other	FBgn0028988	Spn1	effector
immunity	FBgn0019809	gcm2	signaling	other	FBgn0028987	Spn2	effector
immunity	FBgn0010303	hep	signaling	other	FBgn0028986	Spn3	effector
immunity	FBgn0004864	hop	signaling	other	FBgn0028985	Spn4	effector
immunity	FBgn0013983	imd	signaling	other	FBgn0028984	Spn5	effector
immunity	FBgn0024222	ird5	signaling	other	FBgn0038412	Zip3	effector
immunity	FBgn0001291	Jra	signaling	other	FBgn0000210	br	hematopoiesis
immunity	FBgn0001297	kay	signaling	other	FBgn0000463	Dl	hematopoiesis
immunity	FBgn0041205	key	signaling	other	FBgn0001981	esg	hematopoiesis
immunity	FBgn0001319	kn	signaling	other	FBgn0040206	krz	hematopoiesis
immunity	FBgn0015763	lic	signaling	other	FBgn0002440	l(3)mbn	hematopoiesis
immunity	FBgn0010602	lwr	signaling	other	FBgn0015544	spag	hematopoiesis
immunity	FBgn0002576	lz	signaling	other	FBgn0017645	Ank2	insecticide
immunity	FBgn0026207	mbo	signaling	other	FBgn0000120	Arr1	insecticide
immunity	FBgn0024329	Mekk1	signaling	other	FBgn0000121	Arr2	insecticide
immunity	FBgn0024326	Mkk4	signaling	other	FBgn0045761	CG10618	insecticide
immunity	FBgn0015765	Mpk2	signaling	other	FBgn0030484	CG1681	insecticide

Table B.1 (continued)

Category	FBgn	Symbol	Function	Category	FBgn	Symbol	Function
immunity	FBgn0010909	msn	signaling	other	FBgn0027590	CG16936	insecticide
immunity	FBgn0015770	MstProx	signaling	other	FBgn0035113	CG17142	insecticide
immunity	FBgn0005639	mxc	signaling	other	FBgn0050019	CG30019	insecticide
immunity	FBgn0033402	Myd88	signaling	other	FBgn0038878	CG3301	insecticide
immunity	FBgn0004647	N	signaling	other	FBgn0034354	CG5224	insecticide
immunity	FBgn0002930	nec	signaling	other	FBgn0031327	CG5397	insecticide
immunity	FBgn0011676	Nos	signaling	other	FBgn0034888	CG5431	insecticide
immunity	FBgn0031145	Ntf-2	signaling	other	FBgn0035906	CG6673	insecticide
immunity	FBgn0010660	Nup214	signaling	other	FBgn0033728	CG8505	insecticide
immunity	FBgn0004956	os	signaling	other	FBgn0033679	CG8888	insecticide
immunity	FBgn0024846	p38b	signaling	other	FBgn0034807	CG9897	insecticide
immunity	FBgn0003079	phl	signaling	other	FBgn0038681	Cyp12a4	insecticide
immunity	FBgn0010441	pll	signaling	other	FBgn0053503	Cyp12d1-d	insecticide
immunity	FBgn0003118	pnt	signaling	other	FBgn0050489	Cyp12d1-p	insecticide
immunity	FBgn0030926	psh	signaling	other	FBgn0033982	Cyp317a1	insecticide
immunity	FBgn0004210	puc	signaling	other	FBgn0034053	Cyp4aa1	insecticide
immunity	FBgn0030964	Pvf1	signaling	other	FBgn0015037	Cyp4p1	insecticide
immunity	FBgn0031888	Pvf2	signaling	other	FBgn0033302	Cyp6a14	insecticide
immunity	FBgn0031889	Pvf3	signaling	other	FBgn0015714	Cyp6a17	insecticide
immunity	FBgn0032006	Pvr	signaling	other	FBgn0033979	Cyp6a19	insecticide
immunity	FBgn0010333	Rac1	signaling	other	FBgn0000473	Cyp6a2	insecticide
immunity	FBgn0003205	Ras85D	signaling	other	FBgn0033980	Cyp6a20	insecticide
immunity	FBgn0003231	ref(2)P	signaling	other	FBgn0033981	Cyp6a21	insecticide
immunity	FBgn0014018	Rel	signaling	other	FBgn0013773	Cyp6a22	insecticide
immunity	FBgn0004922	RpS6	signaling	other	FBgn0033978	Cyp6a23	insecticide
immunity	FBgn0004197	Ser	signaling	other	FBgn0013772	Cyp6a8	insecticide
immunity	FBgn0030018	slpr	signaling	other	FBgn0013771	Cyp6a9	insecticide
immunity	FBgn0026170	smt3	signaling	other	FBgn0025454	Cyp6g1	insecticide
immunity	FBgn0028990	Spn27A	signaling	other	FBgn0033696	Cyp6g2	insecticide
immunity	FBgn0003495	spz	signaling	other	FBgn0033697	Cyp6t3	insecticide
immunity	FBgn0003507	srp	signaling	other	FBgn0033065	Cyp6w1	insecticide
immunity	FBgn0016917	Stat92E	signaling	other	FBgn0015040	Cyp9c1	insecticide
immunity	FBgn0004837	Su(H)	signaling	other	FBgn0010387	Dbi	insecticide
immunity	FBgn0026323	Tak1	signaling	other	FBgn0011202	dia	insecticide
immunity	FBgn0041582	tamo	signaling	other	FBgn0029175	Ext2	insecticide
immunity	FBgn0026760	Tehao	signaling	other	FBgn0028433	Ggamma30A	insecticide
immunity	FBgn0022073	Thor	signaling	other	FBgn0027657	glob1	insecticide
immunity	FBgn0003717	Tl	signaling	other	FBgn0034334	GstE10	insecticide
immunity	FBgn0032095	Toll-4	signaling	other	FBgn0034340	GstE6	insecticide
immunity	FBgn0036494	Toll-6	signaling	other	FBgn0034343	GstE9	insecticide
immunity	FBgn0034476	Toll-7	signaling	other	FBgn0010226	GstS1	insecticide
immunity	FBgn0036978	Toll-9	signaling	other	FBgn0004784	inaC	insecticide
immunity	FBgn0029114	Tollo	signaling	other	FBgn0023077	inaF	insecticide
immunity	FBgn0026319	Traf1	signaling	other	FBgn0002645	Map205	insecticide
immunity	FBgn0026318	Traf2	signaling	other	FBgn0003036	para	insecticide
immunity	FBgn0030748	Traf3	signaling	other	FBgn0004244	Rdl	insecticide
immunity	FBgn0003882	tub	signaling	other	FBgn0003250	Rh4	insecticide
immunity	FBgn0035601	Uev1A	signaling	other	FBgn0026314	Ugt35b	insecticide
immunity	FBgn0030904	upd2	signaling	other	FBgn0040256	Ugt86Dd	insecticide
immunity	FBgn0003963	ush	signaling	other	FBgn0040252	Ugt86Dh	insecticide
immunity	FBgn0021895	ytr	signaling	other	FBgn0000251	cad	local response
metabolism	FBgn0012034	AcCoAS	\N	other	FBgn0026252	msk	local response
metabolism	FBgn0000024	Ace	\N	other	FBgn0020510	Abi	phagocytosis
metabolism	FBgn0010100	Acon	\N	other	FBgn0000042	Act5C	phagocytosis
metabolism	FBgn0025115	Acyp	\N	other	FBgn0000047	Act88F	phagocytosis
metabolism	FBgn0000055	Adh	\N	other	FBgn0000083	AnnIX	phagocytosis
metabolism	FBgn0010379	Akt1	\N	other	FBgn0010380	Bap	phagocytosis
metabolism	FBgn0000064	Ald	\N	other	FBgn0025724	beta'Cop	phagocytosis
metabolism	FBgn0038742	Arc42	\N	other	FBgn0034179	CG6805	phagocytosis
metabolism	FBgn0013749	Arf102F	\N	other	FBgn0030993	CG7635	phagocytosis
metabolism	FBgn0020236	ATPCL	\N	other	FBgn0036043	CG8177	phagocytosis
metabolism	FBgn0027348	bgm	\N	other	FBgn0000319	Chc	phagocytosis
metabolism	FBgn0041342	Cct1	\N	other	FBgn0000308	chic	phagocytosis
metabolism	FBgn0035231	Cct2	\N	other	FBgn0010434	cora	phagocytosis
metabolism	FBgn0010350	CdsA	\N	other	FBgn0015926	dah	phagocytosis
metabolism	FBgn0037440	CG1041	\N	other	FBgn0028969	deltaCOP	phagocytosis
metabolism	FBgn0035679	CG10467	\N	other	FBgn0026721	fat-spondin	phagocytosis
metabolism	FBgn0029969	CG10932	\N	other	FBgn0038475	Keap1	phagocytosis
metabolism	FBgn0033246	CG11198	\N	other	FBgn0011653	mas	phagocytosis
metabolism	FBgn0037643	CG11963	\N	other	FBgn0020269	mspo	phagocytosis
metabolism	FBgn0037386	CG1208	\N	other	FBgn0015624	nej	phagocytosis
metabolism	FBgn0037356	CG12170	\N	other	FBgn0036101	ninA	phagocytosis
metabolism	FBgn0035811	CG12262	\N	other	FBgn0033247	Nup44A	phagocytosis
metabolism	FBgn0033856	CG13334	\N	other	FBgn0044826	Pak3	phagocytosis

Table B.1 (continued)

Category	FBgn	Symbol	Function	Category	FBgn	Symbol	Function
metabolism	FBgn0037988	CG14740	\N	other	FBgn0020255	ran	phagocytosis
metabolism	FBgn0027580	CG1516	\N	other	FBgn0041781	SCAR	phagocytosis
metabolism	FBgn0037762	CG16905	\N	other	FBgn0013733	shot	phagocytosis
metabolism	FBgn0030239	CG17333	\N	other	FBgn0003475	spir	phagocytosis
metabolism	FBgn0039856	CG1774	\N	other	FBgn0024273	WASp	phagocytosis
metabolism	FBgn0038973	CG18594	\N	other	FBgn0000075	amd	phenoloxidase
metabolism	FBgn0032955	CG2201	\N	other	FBgn0039759	CG9733	phenoloxidase
metabolism	FBgn0051075	CG31075	\N	other	FBgn0037396	CG11459	protease
metabolism	FBgn0027571	CG3523	\N	other	FBgn0039629	CG11842	protease
metabolism	FBgn0032114	CG3752	\N	other	FBgn0038250	CG3505	protease
metabolism	FBgn0029890	CG4095	\N	other	FBgn0032638	CG6639	protease
metabolism	FBgn0028479	CG4389	\N	other	FBgn0038299	CG6687	protease inhibitor
metabolism	FBgn0035006	CG4563	\N	other	FBgn0031973	CG7219	protease inhibitor
metabolism	FBgn0036784	CG5103	\N	other	FBgn0015737	Hmu	recognition
metabolism	FBgn0031912	CG5261	\N	other	FBgn0043364	cabut	signaling
metabolism	FBgn0032237	CG5362	\N	other	FBgn0031114	cactin	signaling
metabolism	FBgn0026576	CG5991	\N	other	FBgn0030909	CG15062	signaling
metabolism	FBgn0033844	CG6016	\N	other	FBgn0030910	CG5963	signaling
metabolism	FBgn0036182	CG6084	\N	other	FBgn0033483	egr	signaling
metabolism	FBgn0039476	CG6271	\N	other	FBgn0028633	ik2	signaling
metabolism	FBgn0029689	CG6428	\N	other	FBgn0040513	kappaB-Ras	signaling
metabolism	FBgn0039184	CG6432	\N	other	FBgn0034118	Nup62	signaling
metabolism	FBgn0033879	CG6543	\N	other	FBgn0014001	Pak	signaling
metabolism	FBgn0038293	CG6904	\N	other	FBgn0025574	Pli	signaling
metabolism	FBgn0031948	CG7149	\N	other	FBgn0003256	rl	signaling
metabolism	FBgn0036762	CG7430	\N	other	FBgn0026178	scrib	signaling
metabolism	FBgn0036691	CG7842	\N	other	FBgn0040271	Sulf1	signaling
metabolism	FBgn0038587	CG7998	\N	other	FBgn0030941	wgn	signaling
metabolism	FBgn0037607	CG8036	\N	other	FBgn0000556	Ef1alpha48D	wounding
metabolism	FBgn0035203	CG9149	\N	other	FBgn0000639	Fbp1	wounding
metabolism	FBgn0034618	CG9485	\N	other	FBgn0002564	Lsp1gamma	wounding
metabolism	FBgn0036857	CG9629	\N	reproduction	FBgn0027080	Aats-tyr	\N
metabolism	FBgn0000303	Cha	\N	reproduction	FBgn0002855	Acp26Aa	\N
metabolism	FBgn0024248	chico	\N	reproduction	FBgn0002856	Acp26Ab	\N
metabolism	FBgn0027842	CPTI	\N	reproduction	FBgn0015583	Acp29AB	\N
metabolism	FBgn0043044	desat1	\N	reproduction	FBgn0023415	Acp32CD	\N
metabolism	FBgn0004568	Dgk	\N	reproduction	FBgn0023414	Acp33A	\N
metabolism	FBgn0020930	Dgkepsilon	\N	reproduction	FBgn0011559	Acp36DE	\N
metabolism	FBgn0040212	Dhap-at	\N	reproduction	FBgn0034152	Acp53C14a	\N
metabolism	FBgn0000472	dm	\N	reproduction	FBgn0034153	Acp53C14b	\N
metabolism	FBgn0000536	eas	\N	reproduction	FBgn0015584	Acp53Ea	\N
metabolism	FBgn0000579	Eno	\N	reproduction	FBgn0020509	Acp62F	\N
metabolism	FBgn0029172	Fad2	\N	reproduction	FBgn0015585	Acp63F	\N
metabolism	FBgn0020440	Fak56D	\N	reproduction	FBgn0003034	Acp70A	\N
metabolism	FBgn0032820	fbp	\N	reproduction	FBgn0015586	Acp76A	\N
metabolism	FBgn0038197	foxo	\N	reproduction	FBgn0002863	Acp95EF	\N
metabolism	FBgn0001075	ft	\N	reproduction	FBgn0013745	Acp98AB	\N
metabolism	FBgn0026718	fu12	\N	reproduction	FBgn0003863	alphaTry	\N
metabolism	FBgn0001091	Gapdh1	\N	reproduction	FBgn0003884	alphaTub84B	\N
metabolism	FBgn0001092	Gapdh2	\N	reproduction	FBgn0012037	Ance	\N
metabolism	FBgn0005198	gig	\N	reproduction	FBgn0016123	Aph-4	\N
metabolism	FBgn0004507	GlyP	\N	reproduction	FBgn0019644	ATPsyn-b	\N
metabolism	FBgn0001124	Got1	\N	reproduction	FBgn0016120	ATPsyn-d	\N
metabolism	FBgn0001125	Got2	\N	reproduction	FBgn0011211	blw	\N
metabolism	FBgn0001128	Gpdh	\N	reproduction	FBgn0011723	byn	\N
metabolism	FBgn0025592	Gyk	\N	reproduction	FBgn0004781	Ccp84Ac	\N
metabolism	FBgn0001186	Hex-A	\N	reproduction	FBgn0035294	CG1017	\N
metabolism	FBgn0001187	Hex-C	\N	reproduction	FBgn0037441	CG10284	\N
metabolism	FBgn0015234	HLH106	\N	reproduction	FBgn0034638	CG10433	\N
metabolism	FBgn0001205	Hmgcr	\N	reproduction	FBgn0032833	CG10664	\N
metabolism	FBgn0010611	Hmgs	\N	reproduction	FBgn0031865	CG10806	\N
metabolism	FBgn0001248	Idh	\N	reproduction	FBgn0034195	CG10956	\N
metabolism	FBgn0044051	Ilp1	\N	reproduction	FBgn0030520	CG10990	\N
metabolism	FBgn0036046	Ilp2	\N	reproduction	FBgn0038067	CG11598	\N
metabolism	FBgn0044050	Ilp3	\N	reproduction	FBgn0040341	CG11664	\N
metabolism	FBgn0044049	Ilp4	\N	reproduction	FBgn0028944	CG11864	\N
metabolism	FBgn0044048	Ilp5	\N	reproduction	FBgn0039313	CG11892	\N
metabolism	FBgn0044047	Ilp6	\N	reproduction	FBgn0025618	CG12311	\N
metabolism	FBgn0044046	Ilp7	\N	reproduction	FBgn0035373	CG1241	\N
metabolism	FBgn0001258	ImpL3	\N	reproduction	FBgn0037213	CG12581	\N
metabolism	FBgn0036816	Indy	\N	reproduction	FBgn0035545	CG12607	\N
metabolism	FBgn0038770	Indy-2	\N	reproduction	FBgn0032789	CG13083	\N
metabolism	FBgn0013984	InR	\N	reproduction	FBgn0035933	CG13309	\N
metabolism	FBgn0034589	king-tubby	\N	reproduction	FBgn0037627	CG13318	\N

Table B.1 (continued)

Category	FBgn	Symbol	Function	Category	FBgn	Symbol	Function
metabolism	FBgn0026708	l(1)G0030	\N	reproduction	FBgn0031691	CG14034	\N
metabolism	FBgn0027291	l(1)G0156	\N	reproduction	FBgn0038147	CG14375	\N
metabolism	FBgn0028336	l(1)G0255	\N	reproduction	FBgn0033046	CG14470	\N
metabolism	FBgn0028325	l(1)G0334	\N	reproduction	FBgn0037928	CG14713	\N
metabolism	FBgn0010609	l(2)44DEa	\N	reproduction	FBgn0030469	CG15745	\N
metabolism	FBgn0022160	l(2)k05713	\N	reproduction	FBgn0037763	CG16904	\N
metabolism	FBgn0023496	Lip1	\N	reproduction	FBgn0032275	CG17097	\N
metabolism	FBgn0023495	Lip3	\N	reproduction	FBgn0032281	CG17107	\N
metabolism	FBgn0029155	Mdh	\N	reproduction	FBgn0032285	CG17108	\N
metabolism	FBgn0004797	mdy	\N	reproduction	FBgn0034877	CG17280	\N
metabolism	FBgn0002719	Men	\N	reproduction	FBgn0033776	CG17575	\N
metabolism	FBgn0011361	mtacp1	\N	reproduction	FBgn0038919	CG17843	\N
metabolism	FBgn0010352	Nc73EF	\N	reproduction	FBgn0032274	CG18284	\N
metabolism	FBgn0032482	Pect	\N	reproduction	FBgn0038347	CG18522	\N
metabolism	FBgn0003067	Pepck	\N	reproduction	FBgn0033241	CG2915	\N
metabolism	FBgn0003071	Pfk	\N	reproduction	FBgn0040499	CG30171	\N
metabolism	FBgn0004654	Pgd	\N	reproduction	FBgn0050404	CG30404	\N
metabolism	FBgn0003074	Pgi	\N	reproduction	FBgn0029804	CG3097	\N
metabolism	FBgn0003075	Pgk	\N	reproduction	FBgn0051163	CG31163	\N
metabolism	FBgn0014869	Pglym78	\N	reproduction	FBgn0051199	CG31199	\N
metabolism	FBgn0003076	Pgm	\N	reproduction	FBgn0051200	CG31200	\N
metabolism	FBgn0033075	Pld	\N	reproduction	FBgn0051645	CG31645	\N
metabolism	FBgn0013955	PR2	\N	reproduction	FBgn0051872	CG31872	\N
metabolism	FBgn0026379	Pten	\N	reproduction	FBgn0051999	CG31999	\N
metabolism	FBgn0020385	pug	\N	reproduction	FBgn0031436	CG3214	\N
metabolism	FBgn0003178	PyK	\N	reproduction	FBgn0052473	CG32473	\N
metabolism	FBgn0003217	rdgA	\N	reproduction	FBgn0052521	CG32521	\N
metabolism	FBgn0016724	RfaBp	\N	reproduction	FBgn0052642	CG32642	\N
metabolism	FBgn0037105	S1P	\N	reproduction	FBgn0052952	CG32952	\N
metabolism	FBgn0033656	S2P	\N	reproduction	FBgn0053171	CG33171	\N
metabolism	FBgn0015806	S6k	\N	reproduction	FBgn0038463	CG3534	\N
metabolism	FBgn0033052	SCAP	\N	reproduction	FBgn0036642	CG4169	\N
metabolism	FBgn0004888	Scsalpha	\N	reproduction	FBgn0038376	CG4225	\N
metabolism	FBgn0017539	Scs-fp	\N	reproduction	FBgn0034132	CG4439	\N
metabolism	FBgn0021765	scu	\N	reproduction	FBgn0029891	CG4523	\N
metabolism	FBgn0014028	SdhB	\N	reproduction	FBgn0038745	CG4538	\N
metabolism	FBgn0024291	Sir2	\N	reproduction	FBgn0034229	CG4847	\N
metabolism	FBgn0030300	Sk1	\N	reproduction	FBgn0038984	CG5315	\N
metabolism	FBgn0052484	Sk2	\N	reproduction	FBgn0032476	CG5439	\N
metabolism	FBgn0033170	sPLA2	\N	reproduction	FBgn0030853	CG5703	\N
metabolism	FBgn0010591	Sply	\N	reproduction	FBgn0038508	CG5866	\N
metabolism	FBgn0029118	Sucb	\N	reproduction	FBgn0039492	CG6051	\N
metabolism	FBgn0033782	sug	\N	reproduction	FBgn0039418	CG6069	\N
metabolism	FBgn0025352	Thiolase	\N	reproduction	FBgn0036154	CG6168	\N
metabolism	FBgn0021796	Tor	\N	reproduction	FBgn0036970	CG6289	\N
metabolism	FBgn0003738	Tpi	\N	reproduction	FBgn0033875	CG6357	\N
metabolism	FBgn0027560	Tps1	\N	reproduction	FBgn0032284	CG7294	\N
metabolism	FBgn0003748	Treh	\N	reproduction	FBgn0032283	CG7296	\N
metabolism	FBgn0004889	tws	\N	reproduction	FBgn0036749	CG7460	\N
metabolism	FBgn0035978	UGP	\N	reproduction	FBgn0037146	CG7470	\N
metabolism	FBgn0042627	v(2)k05816	\N	reproduction	FBgn0040793	CG7630	\N
metabolism	FBgn0016078	wun	\N	reproduction	FBgn0033584	CG7737	\N
metabolism	FBgn0041087	wun2	\N	reproduction	FBgn0030087	CG7766	\N
metabolism	FBgn0040064	yip2	\N	reproduction	FBgn0039737	CG7920	\N
metabolism	FBgn0004057	Zw	\N	reproduction	FBgn0033999	CG8093	\N
other	FBgn0010339	128up	\N	reproduction	FBgn0032945	CG8665	\N
other	FBgn0028550	A3-3	\N	reproduction	FBgn0031745	CG8965	\N
other	FBgn0023129	aay	\N	reproduction	FBgn0031746	CG9029	\N
other	FBgn0014454	Acp1	\N	reproduction	FBgn0034497	CG9090	\N
other	FBgn0020765	Acp65Aa	\N	reproduction	FBgn0036878	CG9283	\N
other	FBgn0026602	Ady43A	\N	reproduction	FBgn0038181	CG9297	\N
other	FBgn0025186	ari-2	\N	reproduction	FBgn0032897	CG9336	\N
other	FBgn0035715	CG10103	\N	reproduction	FBgn0036451	CG9425	\N
other	FBgn0035695	CG10226	\N	reproduction	FBgn0031810	CG9511	\N
other	FBgn0032836	CG10680	\N	reproduction	FBgn0003060	CG9757	\N
other	FBgn0034289	CG10910	\N	reproduction	FBgn0039589	CG9986	\N
other	FBgn0039268	CG11819	\N	reproduction	FBgn0039597	CG9997	\N
other	FBgn0039624	CG11833	\N	reproduction	FBgn0004629	Cys	\N
other	FBgn0039644	CG11897	\N	reproduction	FBgn0024381	Dup99B	\N
other	FBgn0039837	CG12114	\N	reproduction	FBgn0000566	Eip55E	\N
other	FBgn0030041	CG12116	\N	reproduction	FBgn0035849	ERR	\N
other	FBgn0028504	CG12182	\N	reproduction	FBgn0000592	Est-6	\N
other	FBgn0040371	CG12470	\N	reproduction	FBgn0001337	Exp6	\N
other	FBgn0033926	CG12505	\N	reproduction	FBgn0020439	fau	\N

Table B.1 (continued)

Category	FBgn	Symbol	Function	Category	FBgn	Symbol	Function
other	FBgn0035513	CG1259	\N	reproduction	FBgn0011205	fbl	\N
other	FBgn0033521	CG12896	\N	reproduction	FBgn0025519	fidipidine	\N
other	FBgn0033511	CG12907	\N	reproduction	FBgn0020303	fok	\N
other	FBgn0032810	CG13077	\N	reproduction	FBgn0001114	Glt	\N
other	FBgn0032140	CG13117	\N	reproduction	FBgn0001120	gnu	\N
other	FBgn0033788	CG13323	\N	reproduction	FBgn0013972	Gycalpha99B	\N
other	FBgn0033789	CG13324	\N	reproduction	FBgn0042712	HBS1	\N
other	FBgn0033857	CG13335	\N	reproduction	FBgn0014857	His3.3A	\N
other	FBgn0032037	CG13394	\N	reproduction	FBgn0004167	kst	\N
other	FBgn0036419	CG13482	\N	reproduction	FBgn0011296	l(2)efl	\N
other	FBgn0034695	CG13503	\N	reproduction	FBgn0011013	l(3)s1921	\N
other	FBgn0033394	CG13740	\N	reproduction	FBgn0040098	lectin-29Ca	\N
other	FBgn0031937	CG13795	\N	reproduction	FBgn0040093	lectin-46Ca	\N
other	FBgn0035112	CG13877	\N	reproduction	FBgn0040092	lectin-46Cb	\N
other	FBgn0035176	CG13905	\N	reproduction	FBgn0005278	M(2)21AB	\N
other	FBgn0031032	CG14204	\N	reproduction	FBgn0029870	Marf	\N
other	FBgn0031033	CG14219	\N	reproduction	FBgn0024211	mfas	\N
other	FBgn0032900	CG14401	\N	reproduction	FBgn0002741	Mhc	\N
other	FBgn0037131	CG14564	\N	reproduction	FBgn0002772	Mlc1	\N
other	FBgn0037126	CG14567	\N	reproduction	FBgn0002773	Mlc2	\N
other	FBgn0037123	CG14569	\N	reproduction	FBgn0002789	Mp20	\N
other	FBgn0037288	CG14661	\N	reproduction	FBgn0004414	msopa	\N
other	FBgn0035797	CG14837	\N	reproduction	FBgn0011668	Mst57Da	\N
other	FBgn0035412	CG14957	\N	reproduction	FBgn0011669	Mst57Db	\N
other	FBgn0030927	CG15046	\N	reproduction	FBgn0011670	Mst57Dc	\N
other	FBgn0040734	CG15065	\N	reproduction	FBgn0004657	mys	\N
other	FBgn0034331	CG15067	\N	reproduction	FBgn0019957	ND42	\N
other	FBgn0034394	CG15096	\N	reproduction	FBgn0017566	ND75	\N
other	FBgn0040729	CG15126	\N	reproduction	FBgn0005322	nmd	\N
other	FBgn0037398	CG15580	\N	reproduction	FBgn0014366	noi	\N
other	FBgn0034647	CG15678	\N	reproduction	FBgn0020386	Pk61C	\N
other	FBgn0029766	CG15784	\N	reproduction	FBgn0003149	Prm	\N
other	FBgn0032835	CG16772	\N	reproduction	FBgn0033518	Prx2540-2	\N
other	FBgn0037713	CG16790	\N	reproduction	FBgn0004368	Ptp4E	\N
other	FBgn0040735	CG16836	\N	reproduction	FBgn0014009	Rab2	\N
other	FBgn0040972	CG16978	\N	reproduction	FBgn0030986	RhoGAP18B	\N
other	FBgn0039045	CG17119	\N	reproduction	FBgn0010406	RNaseX25	\N
other	FBgn0031488	CG17265	\N	reproduction	FBgn0003279	RpL4	\N
other	FBgn0046763	CG17278	\N	reproduction	FBgn0011834	Ser6	\N
other	FBgn0028394	CG17834	\N	reproduction	FBgn0025803	SNF4Agamma	\N
other	FBgn0034512	CG18067	\N	reproduction	FBgn0003464	sol	\N
other	FBgn0042120	CG18779	\N	reproduction	FBgn0024293	Spn43Ab	\N
other	FBgn0027544	CG2217	\N	reproduction	FBgn0028983	Spn6	\N
other	FBgn0050076	CG30076	\N	reproduction	FBgn0023477	Tal	\N
other	FBgn0051638	CG31638	\N	reproduction	FBgn0023479	Tequila	\N
other	FBgn0052074	CG32074	\N	reproduction	FBgn0004117	Tm2	\N
other	FBgn0052214	CG32214	\N	reproduction	FBgn0039668	Trc8	\N
other	FBgn0052412	CG32412	\N	reproduction	FBgn0001402	trol	\N
other	FBgn0040368	CG32859	\N	reproduction	FBgn0003964	usp	\N
other	FBgn0053047	CG33047	\N	reproduction	FBgn0045823	vsg	\N
other	FBgn0058192	CG40192	\N				

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

Table C.1 Transcribed Genes Classified Among Immune Functional Groups

Effector		Recognition		Signaling	
FBgn	**Symbol**	**FBgn**	**Symbol**	**FBgn**	**Symbol**
FBgn0041579	AttC	FBgn0033301	CG12780	FBgn0029512	Aos1
FBgn0000276	CecA1	FBgn0032808	CG13079	FBgn0000173	ben
FBgn0000277	CecA2	FBgn0034511	CG13422	FBgn0000229	bsk
FBgn0000278	CecB	FBgn0050148	CG30148	FBgn0000250	cact
FBgn0035734	CG14823	FBgn0040322	GNBP2	FBgn0039666	CG11501
FBgn0028526	CG15293	FBgn0028430	He	FBgn0030925	CG6361
FBgn0034330	CG18107	FBgn0054003	nimB3	FBgn0014141	cher
FBgn0025827	CG6421	FBgn0028939	nimC2	FBgn0011274	Dif
FBgn0034092	CG7798	FBgn0035976	PGRP-LC	FBgn0020306	dom
FBgn0000487	Dox-A3	FBgn0030695	PGRP-LE	FBgn0043903	dome
FBgn0004240	Dpt	FBgn0035977	PGRP-LF	FBgn0028436	ECSIT
FBgn0034407	DptB	FBgn0030310	PGRP-SA	FBgn0014179	gcm
FBgn0010388	Dro	FBgn0043575	PGRP-SC2	FBgn0022787	Hel89B
FBgn0052279	dro2	FBgn0014033	Sr-CI	FBgn0010303	hep
FBgn0052283	dro3	FBgn0020376	Sr-CIII	FBgn0041205	key
FBgn0052268	dro6	FBgn0041183	TepI	FBgn0001319	kn
FBgn0010381	Drs	FBgn0041180	TepIV	FBgn0015765	Mpk2
FBgn0029167	Hml			FBgn0031145	Ntf-2
FBgn0034329	IM1			FBgn0032680	Ntf-2r
FBgn0025583	IM2			FBgn0004956	os
FBgn0034328	IM23			FBgn0040294	POSH
FBgn0040736	IM3			FBgn0032006	Pvr
FBgn0004425	LysB			FBgn0003231	ref(2)P
FBgn0004428	LysE			FBgn0026170	smt3
FBgn0004429	LysP			FBgn0041184	Socs36E
FBgn0004431	LysX			FBgn0039102	SPE
FBgn0014865	Mtk			FBgn0052382	sphinx2
FBgn0005626	ple			FBgn0030051	spirit
FBgn0003162	Pu			FBgn0026323	Tak1
FBgn0028396	TotA			FBgn0041582	tamo
FBgn0044812	TotC			FBgn0022073	Thor
FBgn0031701	TotM			FBgn0032095	Toll-4
FBgn0044809	TotZ			FBgn0026319	Traf4
				FBgn0003882	tub
				FBgn0029113	Uba2
				FBgn0035601	Uev1A
				FBgn0027603	Ulp1
				FBgn0038134	wntD

APPENDIX D

SUPPLEMENTARY INFORMATION FOR CHAPTER 5

Table D.1 Mutant *D. melanogaster* stocks used in crosses

Gene	Exelixis Stock	Genetic Background
spz	XP d00069	w^{1118}
MyD88	XP d09821	w^{1118}
Tub	RB e03259	w^{1118}
Pll	RB e01040	w^{1118}
Dif	XP d05062	w^{1118}
Cact	XP d10397	w^{1118}
PGRP-LC	XP d04396	w^{1118}
imd	XP d00066	w^{1118}
BG4	WH f02804	w^{1118}
DREDD	EP 1412	w^{1118}
ird5	DrosGDP*	$y^{1}w^{67c23};ry^{506}$
key	PB c02831	w^{1118}
Rel	XP d03528	w^{1118}

*Dros. Gene Disruption project,
stock Bl-14684

Table D.2 Primer and Probe Sequences for *Dpt* Pyrosequencing

Primer	Sequence
forward - 5' biotinylated	CAGCACTTGGGAGGACCATA
reverse	TGTAGGTGCTTCCCACTTTCC
probe	GGTGCTTCCCACTTT